

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicant(s): T. OKAMOTO et al  
Serial No.: Not Yet Assigned  
Filed: On Even Date Herewith  
Title: DISK ARRAY APPARATUS AND CONTROL METHOD  
FOR THE SAME

**LETTER CLAIMING RIGHT OF PRIORITY**

**Mail Stop: Patent Applications**

May 20, 2004


Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on **Japanese** Patent Application No. **2004-063313**, filed March 8, 2004.

A certified copy of said Japanese Application is attached.

Respectfully submitted,  
**ANTONELLI, TERRY, STOUT & KRAUS, LLP**

  
\_\_\_\_\_  
Carl I. Brundidge  
Registration Number 29,621

CIB/dks  
Attachment  
(703) 312-6600

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日                      2 0 0 4 年    3 月    8 日  
Date of Application:

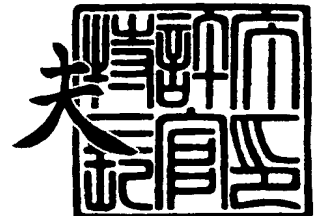
出 願 番 号                      特 願 2 0 0 4 - 0 6 3 3 1 3  
Application Number:  
[ST. 10/C]:                      [ J P 2 0 0 4 - 0 6 3 3 1 3 ]

出      願      人                      株式会社日立製作所  
Applicant(s):

2 0 0 4 年    4 月    7 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



出証番号    出証特 2 0 0 4 - 3 0 2 8 5 6 7

【書類名】 特許願  
【整理番号】 K04001481  
【提出日】 平成16年 3月 8日  
【あて先】 特許庁長官殿  
【国際特許分類】 G06F 3/06  
【発明者】  
    【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A  
                          I D システム事業部内  
    【氏名】 岡本 岳樹  
【発明者】  
    【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A  
                          I D システム事業部内  
    【氏名】 福岡 幹夫  
【特許出願人】  
    【識別番号】 000005108  
    【氏名又は名称】 株式会社日立製作所  
【代理人】  
    【識別番号】 100080001  
    【弁理士】  
    【氏名又は名称】 筒井 大和  
    【電話番号】 03-3366-0787  
【手数料の表示】  
    【予納台帳番号】 006909  
    【納付金額】 21,000円  
【提出物件の目録】  
    【物件名】 特許請求の範囲 1  
    【物件名】 明細書 1  
    【物件名】 図面 1  
    【物件名】 要約書 1

**【書類名】 特許請求の範囲****【請求項 1】**

記憶領域に対するデータの書き込みまたは読み出し処理においてエラーが生じた場合、データの書き込みまたは読み出しが正常に終了したことを通知した後に、データの書き込みまたは読み出しを再度繰り返し、データの書き込みまたは読み出しを実行する複数の記憶デバイスと、

前記複数の記憶デバイスに対する書き込みデータまたは読み出し要求が格納される格納領域を有して、前記複数の記憶デバイスに対するデータの書き込みまたは読み出しを制御し、前記複数の記憶デバイスからデータの書き込みまたは読み出しが正常に終了したことを通知される記憶デバイス制御部と、

自ディスクアレイ装置の外部のネットワークから書き込み要求または読み出し要求を受けるチャネル制御部と、

前記チャネル制御部、前記記憶デバイス制御部によって通信される制御情報が格納される共有メモリと、

前記チャネル制御部と前記記憶デバイス制御部との間で通信されるデータが一時的に保存されるキャッシュメモリと、

前記チャネル制御部、前記記憶デバイス制御部、前記共有メモリおよび前記キャッシュメモリに接続される接続部とを有し、

前記記憶デバイス制御部は、

データの書き込みまたは読み出しに利用され、冗長性を有してデータを格納できる論理的な記憶領域を、前記複数の記憶デバイスの記憶領域を用いて生成し、

前記論理的な記憶領域を形成する複数の記憶デバイスに関して、前記複数の記憶デバイスに対する書き込みデータまたは読み出し要求が格納される格納領域を監視し、

前記論理的な記憶領域を形成する複数の記憶デバイスのうちデータの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定して、前記特定された記憶デバイスを閉塞させることを特徴とするディスクアレイ装置。

**【請求項 2】**

請求項 1 記載のディスクアレイ装置において、

前記複数の記憶デバイスは、冗長性を有し、

前記共有メモリは、前記キャッシュメモリ上の前記記憶デバイスに書き込むべき未反映データ量を保持する領域を有し、

前記記憶デバイス制御部は、前記データの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定する際に、冗長性を有する前記複数の記憶デバイスの未反映データ量を比較し、未反映データ量の多い記憶デバイスを閉塞させる記憶デバイスとして特定することを特徴とするディスクアレイ装置。

**【請求項 3】**

請求項 1 記載のディスクアレイ装置において、

前記複数の記憶デバイスは、冗長性を有し、

前記共有メモリは、前記複数の記憶デバイスのそれぞれの平均応答時間を保持する領域を有し、

前記記憶デバイス制御部は、前記データの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定する際に、冗長性を有する前記複数の記憶デバイスの平均応答時間を比較し、平均応答時間の大きい記憶デバイスを閉塞させる記憶デバイスとして特定することを特徴とするディスクアレイ装置。

**【請求項 4】**

請求項 1 記載のディスクアレイ装置において、

前記複数の記憶デバイスは、冗長性を有し、

前記記憶デバイス制御部は、前記複数の記憶デバイスのそれぞれに対するキュー数を保持する領域を有し、前記データの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定する際に、冗長性を有する前記複数の記憶デバイスのキュー数を比較し、キュー

一数が多い記憶デバイスを閉塞させる記憶デバイスとして特定することを特徴とするディスクアレイ装置。

【請求項 5】

請求項 1 記載のディスクアレイ装置において、  
前記複数の記憶デバイスは、冗長性を有し、  
前記記憶デバイス制御部は、前記データの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定して閉塞させた後、冗長性を有する他の記憶デバイスのデータを用いて予備の記憶デバイスにデータを復元することを特徴とするディスクアレイ装置。

【請求項 6】

請求項 1 記載のディスクアレイ装置において、  
前記複数の記憶デバイスは、冗長性を有し、  
前記接続部に接続される管理端末をさらに有し、  
前記管理端末は、冗長性を有する前記複数の記憶デバイスのうち前記データの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定する際の条件を設定することを特徴とするディスクアレイ装置。

【請求項 7】

請求項 6 記載のディスクアレイ装置において、  
前記条件は、前記データの書き込みまたは読み出しの繰り返し回数であることを特徴とするディスクアレイ装置。

【請求項 8】

請求項 6 記載のディスクアレイ装置において、  
前記条件は、前記キャッシュメモリ上の前記記憶デバイスに書き込むべき未反映データ量の差分倍数であることを特徴とするディスクアレイ装置。

【請求項 9】

請求項 6 記載のディスクアレイ装置において、  
前記条件は、前記複数の記憶デバイスのそれぞれの平均応答時間の差分倍数であることを特徴とするディスクアレイ装置。

【請求項 10】

請求項 6 記載のディスクアレイ装置において、  
前記条件は、前記複数の記憶デバイスのそれぞれに対するキュー数の差分倍数であることを特徴とするディスクアレイ装置。

【請求項 11】

記憶領域に対するデータの書き込みまたは読み出し処理においてエラーが生じた場合、データの書き込みまたは読み出しが正常に終了したことを通知した後に、データの書き込みまたは読み出しを再度繰り返し、データの書き込みまたは読み出しを実行する複数の記憶デバイスと、

前記複数の記憶デバイスに対する書き込みデータまたは読み出し要求が格納される格納領域を有して、前記複数の記憶デバイスに対するデータの書き込みまたは読み出しを制御し、前記複数の記憶デバイスからデータの書き込みまたは読み出しが正常に終了したことを通知される記憶デバイス制御部と、

自ディスクアレイ装置の外部のネットワークから書き込み要求または読み出し要求を受けるチャンネル制御部と、

前記チャンネル制御部、前記記憶デバイス制御部によって通信される制御情報が格納される共有メモリと、

前記チャンネル制御部と前記記憶デバイス制御部との間で通信されるデータが一時的に保存されるキャッシュメモリと、

前記チャンネル制御部、前記記憶デバイス制御部、前記共有メモリおよび前記キャッシュメモリに接続される接続部とを有するディスクアレイ装置の制御方法であって、

前記記憶デバイス制御部は、

データの書き込みまたは読み出しに利用され、冗長性を有してデータを格納できる論

理的な記憶領域を、前記複数の記憶デバイスの記憶領域を用いて生成し、

前記論理的な記憶領域を形成する複数の記憶デバイスに関して、前記複数の記憶デバイスに対する書き込みデータまたは読み出し要求が格納される格納領域を監視し、

前記論理的な記憶領域を形成する複数の記憶デバイスのうちデータの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定して、前記特定された記憶デバイスを閉塞させることを特徴とするディスクアレイ装置の制御方法。

【請求項 1 2】

請求項 1 1 記載のディスクアレイ装置の制御方法において、

前記複数の記憶デバイスは、冗長性を有し、

前記共有メモリは、前記キャッシュメモリ上の前記記憶デバイスに書き込むべき未反映データ量を保持する領域を有し、

前記記憶デバイス制御部は、前記データの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定する際に、冗長性を有する前記複数の記憶デバイスの未反映データ量を比較し、未反映データ量の多い記憶デバイスを閉塞させる記憶デバイスとして特定することを特徴とするディスクアレイ装置の制御方法。

【請求項 1 3】

請求項 1 1 記載のディスクアレイ装置の制御方法において、

前記複数の記憶デバイスは、冗長性を有し、

前記共有メモリは、前記複数の記憶デバイスのそれぞれの平均応答時間を保持する領域を有し、

前記記憶デバイス制御部は、前記データの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定する際に、冗長性を有する前記複数の記憶デバイスの平均応答時間を比較し、平均応答時間の大きい記憶デバイスを閉塞させる記憶デバイスとして特定することを特徴とするディスクアレイ装置の制御方法。

【請求項 1 4】

請求項 1 1 記載のディスクアレイ装置の制御方法において、

前記複数の記憶デバイスは、冗長性を有し、

前記記憶デバイス制御部は、前記複数の記憶デバイスのそれぞれに対するキュー数を保持する領域を有し、前記データの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定する際に、冗長性を有する前記複数の記憶デバイスのキュー数を比較し、キュー数が多い記憶デバイスを閉塞させる記憶デバイスとして特定することを特徴とするディスクアレイ装置の制御方法。

【請求項 1 5】

請求項 1 1 記載のディスクアレイ装置の制御方法において、

前記複数の記憶デバイスは、冗長性を有し、

前記記憶デバイス制御部は、前記データの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定して閉塞させた後、冗長性を有する他の記憶デバイスのデータを用いて予備の記憶デバイスにデータを復元することを特徴とするディスクアレイ装置の制御方法。

【請求項 1 6】

請求項 1 1 記載のディスクアレイ装置の制御方法において、

前記複数の記憶デバイスは、冗長性を有し、

前記接続部に接続される管理端末をさらに有し、

前記管理端末は、冗長性を有する前記複数の記憶デバイスのうち前記データの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定する際の条件を設定することを特徴とするディスクアレイ装置の制御方法。

【請求項 1 7】

請求項 1 6 記載のディスクアレイ装置の制御方法において、

前記条件は、前記データの書き込みまたは読み出しの繰り返し回数であることを特徴とするディスクアレイ装置の制御方法。

**【請求項 18】**

請求項 16 記載のディスクアレイ装置の制御方法において、  
前記条件は、前記キャッシュメモリ上の前記記憶デバイスに書き込むべき未反映データの差分倍数であることを特徴とするディスクアレイ装置の制御方法。

**【請求項 19】**

請求項 16 記載のディスクアレイ装置の制御方法において、  
前記条件は、前記複数の記憶デバイスのそれぞれの平均応答時間の差分倍数であることを特徴とするディスクアレイ装置の制御方法。

**【請求項 20】**

請求項 16 記載のディスクアレイ装置の制御方法において、  
前記条件は、前記複数の記憶デバイスのそれぞれに対するキュー数の差分倍数であることを特徴とするディスクアレイ装置の制御方法。

【書類名】 明細書

【発明の名称】 ディスクアレイ装置およびその制御方法

【技術分野】

【0001】

本発明は、ディスクアレイ装置およびその制御技術に関し、特に、性能劣化しているディスクドライブの検出に適用して有効な技術に関する。

【背景技術】

【0002】

本発明者が検討したところによれば、従来のディスクアレイ装置およびその制御技術に関して、以下のような技術が考えられる。

【0003】

たとえば、従来のディスクアレイ装置およびその制御技術においては、ディスクドライブに対するデータの書き込みまたは読み出し処理においてエラーが生じた場合、データの書き込みまたは読み出しを繰り返して行う処理、いわゆるリトライが行われる。このリトライにおいては、発生したエラー要因を取得し、この取得したエラー要因に対応するリトライ回数を設定し、この回数以内で救済に成功した場合にはリカバードとしてホストに報告される（特許文献1参照）。

【0004】

ところで、ディスクアレイ装置のディスクドライブは、年々大容量化に伴い記録密度も高密度化の一途を辿っている。そのため、記録密度当たりのエラーレートは、従来のディスクドライブと同等であっても、ディスクドライブ当たりでのエラー件数は記録密度に比例することになり、従来と同じようにリトライにより救済できた事例を毎々ホストに報告してしまうと、エラー報告件数が多大となり、すぐに故障ディスクドライブと判断されやすくなる。

【0005】

その打開策の一例として、たとえばリトライ回数が数回であった場合には、ホストにリカバード（救済成功）ではなく、正常終了を報告するような方式、いわゆる隠しリトライを採用することがある。この隠しリトライにおいて、ディスクドライブを使用するホストは、通常、ディスクドライブの応答時間を監視しているが、突発的な一時障害を考慮し、監視時間は正常な応答時間の10倍～100倍程度に設定していることが常である。そのため、慢性的に救済処理を行い、応答時間が常に正常時の1.5～2倍程度に遅延しているディスクドライブを遅延と検出することはない。

【特許文献1】 特開2003-141824号公報

【発明の開示】

【発明が解決しようとする課題】

【0006】

ところで、前記のような本発明者が検討した従来のディスクアレイ装置およびその制御技術に関して、本発明者が検討した結果、以下のようなことが明らかとなった。

【0007】

たとえば、従来のディスクアレイ装置およびその制御技術においては、前述の通り、ディスクドライブはホストには見えない隠しリトライや、またはより複雑になったヘッド制御シーケンスの $\mu$ 不良等により、ホストに正常終了と報告していても、その報告時間は正常時の数倍になることがあり、それが慢性的に発生すれば、慢性的な性能劣化ドライブとなる。このように、ホスト側が単純に正常時の報告時間の数倍を目処に監視時間を設定してしまうと、突発的な一時障害も検出してしまうため、これを区別するためには統計的に判断する必要がある。

【0008】

そこで、本発明の目的は、ディスクドライブ自身に性能劣化を検出する機能がなくても、性能劣化しているディスクドライブを検出することができ、さらに性能劣化レベルを可変できることにより、顧客要求に応じたディスクアレイ装置の構築を実現することができ



る技術を提供することにある。

【0009】

本発明の前記ならびにその他の目的と新規な特徴は、本明細書の記述および添付図面から明らかになるであろう。

【課題を解決するための手段】

【0010】

本願において開示される発明のうち、代表的なものの概要を簡単に説明すれば、次のとおりである。

【0011】

本発明は、記憶領域に対するデータの書き込みまたは読み出し処理においてエラーが生じた場合、データの書き込みまたは読み出しが正常に終了したことを通知した後に、データの書き込みまたは読み出しを再度繰り返し、データの書き込みまたは読み出しを実行する複数の記憶デバイスと、複数の記憶デバイスに対する書き込みデータまたは読み出し要求が格納される格納領域を有して、複数の記憶デバイスに対するデータの書き込みまたは読み出しを制御し、複数の記憶デバイスからデータの書き込みまたは読み出しが正常に終了したことを通知される記憶デバイス制御部と、自ディスクアレイ装置の外部のネットワークから書き込み要求または読み出し要求を受けるチャンネル制御部と、チャンネル制御部、記憶デバイス制御部によって通信される制御情報が格納される共有メモリと、チャンネル制御部と記憶デバイス制御部との間で通信されるデータが一時的に保存されるキャッシュメモリと、チャンネル制御部、記憶デバイス制御部、共有メモリおよびキャッシュメモリに接続される接続部とを有するディスクアレイ装置およびその制御方法に適用され、以下のような特徴を有するものである。

【0012】

すなわち、本発明のディスクアレイ装置において、記憶デバイス制御部は、データの書き込みまたは読み出しに利用され、冗長性を有してデータを格納できる論理的な記憶領域を、複数の記憶デバイスの記憶領域を用いて生成し、論理的な記憶領域を形成する複数の記憶デバイスに関して、複数の記憶デバイスに対する書き込みデータまたは読み出し要求が格納される格納領域を監視し、論理的な記憶領域を形成する複数の記憶デバイスのうちデータの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定して、特定された記憶デバイスを閉塞させるものである。

【0013】

具体的には、データの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定する際に、冗長性を有する複数の記憶デバイスの未反映データ量を比較し、未反映データ量の多い記憶デバイスを閉塞させる記憶デバイスとして特定したり、または冗長性を有する複数の記憶デバイスの平均応答時間を比較し、平均応答時間の大きい記憶デバイスを閉塞させる記憶デバイスとして特定したり、または冗長性を有する複数の記憶デバイスのキュー数を比較し、キュー数が多い記憶デバイスを閉塞させる記憶デバイスとして特定するものである。

【0014】

さらに、接続部に接続される管理端末を有し、冗長性を有する複数の記憶デバイスのうちデータの書き込みまたは読み出しの繰り返し回数が多い記憶デバイスを特定する際の条件、たとえばデータの書き込みまたは読み出しの繰り返し回数、キャッシュメモリ上の記憶デバイスに書き込むべき未反映データ量の差分倍数、複数の記憶デバイスのそれぞれの平均応答時間の差分倍数、複数の記憶デバイスのそれぞれに対するキュー数の差分倍数、などを設定できるようにしたものである。

【発明の効果】

【0015】

本願において開示される発明のうち、代表的なものによって得られる効果を簡単に説明すれば以下のとおりである。

【0016】

本発明によれば、ディスクドライブ自身に性能劣化を検出する機能がなくても、性能劣化しているディスクドライブを検出することが可能であり、さらに性能劣化レベルを可変できることにより、顧客要求に応じたディスクアレイ装置の構築を実現することが可能となる。

【発明を実施するための最良の形態】

【0017】

以下、本発明の実施の形態を図面に基づいて詳細に説明する。なお、実施の形態を説明するための全図において、同一の機能を有する部材には原則として同一の符号を付し、その繰り返しの説明は省略する。

【0018】

＜本発明の概念＞

本発明は、記憶領域に対するデータの書き込みまたは読み出し処理においてエラーが生じた場合、データの書き込みまたは読み出しが正常に終了したことを通知した後に、データの書き込みまたは読み出しを再度繰り返し、データの書き込みまたは読み出しを実行する複数のディスクドライブ（記憶デバイス）と、複数のディスクドライブに対する書き込みデータまたは読み出し要求が格納される格納領域を有して、複数のディスクドライブに対するデータの書き込みまたは読み出しを制御し、複数のディスクドライブからデータの書き込みまたは読み出しが正常に終了したことを通知されるディスクコントローラ（記憶デバイス制御部）と、自ディスクアレイ装置の外部のネットワークから書き込み要求または読み出し要求を受けるチャンネルコントローラ（チャンネル制御部）と、チャンネルコントローラ、ディスクコントローラによって通信される制御情報が格納される共有メモリと、チャンネルコントローラとディスクコントローラとの間で通信されるデータが一時的に保存されるキャッシュメモリと、チャンネルコントローラ、ディスクコントローラ、共有メモリおよびキャッシュメモリに接続されるスイッチ（接続部）と、スイッチに接続されるサービスプロセッサ（管理端末）を有するディスクアレイ装置に適用される。

【0019】

＜ディスクアレイ装置の全体構成例＞

図1により、本発明の一実施の形態に係るディスクアレイ装置の全体構成の一例を説明する。図1は本実施の形態に係るディスクアレイ装置の全体構成を示す構成図である。

【0020】

本実施の形態に係るディスクアレイ装置100は、ディスク制御装置200とディスク駆動装置300とを備える。ディスク制御装置200は、たとえばホストの情報処理装置400から受信したコマンドに従ってディスク駆動装置300に対する制御を行う。たとえば、ホストの情報処理装置400からデータ入出力要求を受信して、ディスク駆動装置300が備えるディスクドライブ310に記憶されるデータの読み書きを行う。また、ディスク制御装置200は、たとえば管理クライアントの情報処理装置500からディスクアレイ装置100を管理するための各種コマンドを受信して、ディスクアレイ装置100の様々な設定を行う。

【0021】

ホスト、管理クライアントの情報処理装置400、500は、CPUやメモリを備えたコンピュータなどの情報機器である。情報処理装置400、500が備えるCPUによって各種プログラムが実行されることにより、様々な機能が実現される。情報処理装置400、500は、たとえばパーソナルコンピュータやワークステーションであることもあるし、メインフレームコンピュータであることもある。特に、ホストの情報処理装置400は、たとえば銀行の自動預金預け払いシステムや航空機の座席予約システムなどにおける中枢コンピュータとして利用される。また、管理クライアントの情報処理装置500は、ディスクアレイ装置100を保守、管理するための管理コンピュータとして利用される。

【0022】

ここで、ホスト、管理クライアントの情報処理装置400、500は、異なるユーザの情報処理装置とすることができる。たとえば、ホストの情報処理装置(1)、(2)40

0、管理クライアントの情報処理装置(6)500は、ユーザAの情報処理装置、ホストの情報処理装置(3)~(5)400、管理クライアントの情報処理装置(7)500は、ユーザBの情報処理装置とすることができる。また、管理クライアントの情報処理装置(8)500は、ディスクアレイ装置100の全体を管理する管理者の情報処理装置とすることができる。ここで、ユーザとは、たとえば企業とすることができ、または企業内における部署などの部門とすることもでき、あるいは個人とすることもできる。

#### 【0023】

図1において、ホストの情報処理装置400は、SAN(Storage Area Network)600を介してディスク制御装置200と通信可能に接続されている。SAN600は、ディスク駆動装置300が提供する記憶資源におけるデータの管理単位であるブロックを単位としてホストの情報処理装置400との間でデータの授受を行うためのネットワークである。SAN600を介して行われるホストの情報処理装置400とディスク制御装置200との間の通信は、たとえばファイバチャネルプロトコルに従って行われるようにすることができる。

#### 【0024】

もちろん、ホストの情報処理装置400とディスク制御装置200との間は、SAN600を介して接続されている必要はなく、たとえば、LAN(Local Area Network)を介して接続されているようにすることもできるし、ネットワークを介さずに直接に接続されているようにすることもできる。LANを介して接続される場合には、たとえばTCP/IP(Transmission Control Protocol/Internet Protocol)プロトコルに従って通信を行うようにすることができる。また、ネットワークを介さずに直接に接続される場合には、たとえばFICON(Fibre Connection)(登録商標)やESCON(Enterprise System Connection)(登録商標)、ACONARC(Advanced Connection Architecture)(登録商標)、FIBARC(Fibre Connection Architecture)(登録商標)などの通信プロトコルに従って通信を行うようにすることもできる。

#### 【0025】

また、管理クライアントの情報処理装置500は、LAN700を介してディスク制御装置200と接続されている。LAN700は、インターネットとすることもできるし、専用のネットワークとすることもできる。LAN700を介して行われる管理クライアントの情報処理装置500とディスク制御装置200との間の通信は、たとえばTCP/IPプロトコルに従って行われるようにすることができる。

#### 【0026】

##### <<ディスク駆動装置>>

ディスク駆動装置300は、多数のディスクドライブ310を備えている。これにより、ホストの情報処理装置400、管理クライアントの情報処理装置500に対して大容量の記憶領域を提供することができる。ディスクドライブ310は、ハードディスクドライブなどのデータ記憶媒体、あるいはRAID(Redundant Arrays of Inexpensive Disks)を構成する複数のハードディスクドライブにより構成されてなるようにすることができる。また、ディスクドライブ310により提供される物理的な記憶領域である物理ボリュームには、論理的な記録領域である論理ボリュームを設定することができる。

#### 【0027】

ディスク制御装置200とディスク駆動装置300との間は、図1のように直接に接続される形態とすることもできるし、ネットワークを介して接続されるようにすることもできる。さらに、ディスク駆動装置300は、ディスク制御装置200と一体として構成されることもできる。

#### 【0028】

##### <<ディスク制御装置>>

ディスク制御装置 200 は、チャンネルコントローラ 210、共有メモリ 220、キャッシュメモリ 230、ディスクコントローラ 240、サービスプロセッサ 250、スイッチ 260 を備える。ディスク制御装置 200 は、チャンネルコントローラ 210 により SAN 600 を介してホストの情報処理装置 400 との間の通信を行う。

#### 【0029】

チャンネルコントローラ 210 は、ホストの情報処理装置 400、管理クライアントの情報処理装置 500 との間で通信を行うための通信インタフェースを備え、ホスト、管理クライアントの情報処理装置 400、500 との間でデータ入出力コマンドなどを授受する機能を備える。

#### 【0030】

各チャンネルコントローラ 210 は、サービスプロセッサ 250 と共に内部 LAN 261 で接続されている。これにより、チャンネルコントローラ 210 に実行させるマイクロプログラムなどをサービスプロセッサ 250 から送信し、インストールすることが可能となっている。

#### 【0031】

スイッチ 260 は、チャンネルコントローラ 210、共有メモリ 220、キャッシュメモリ 230、ディスクコントローラ 240、サービスプロセッサ 250 を相互に接続する。チャンネルコントローラ 210、共有メモリ 220、キャッシュメモリ 230、ディスクコントローラ 240、サービスプロセッサ 250 の間でのデータやコマンドの授受は、スイッチ 260 を介することにより行われる。スイッチ 260 は、たとえばクロスバスイッチで構成される。

#### 【0032】

共有メモリ 220 およびキャッシュメモリ 230 は、チャンネルコントローラ 210、ディスクコントローラ 240 により共有される記憶メモリである。共有メモリ 220 は、主に制御情報やコマンドなどを記憶するために利用されるのに対して、キャッシュメモリ 230 は、主にデータを記憶するために利用される。

#### 【0033】

たとえば、あるチャンネルコントローラ 210 がホストの情報処理装置 400 から受信したデータ入出力要求が書き込みコマンドであった場合には、当該チャンネルコントローラ 210 は書き込みコマンドを共有メモリ 220 に書き込むと共に、ホストの情報処理装置 400 から受信した書き込みデータをキャッシュメモリ 230 に書き込む。一方、ディスクコントローラ 240 は、共有メモリ 220 を監視しており、共有メモリ 220 に書き込みコマンドが書き込まれたことを検出すると、当該コマンドに従ってキャッシュメモリ 230 から書き込みデータを読み出してディスク駆動装置 300 内のディスクドライブ 310 に書き込む。

#### 【0034】

また、あるチャンネルコントローラ 210 がホストの情報処理装置 400 から受信したデータ入出力要求が読み出しコマンドであった場合には、読み出し対象となるデータがキャッシュメモリ 230 に存在するかどうかを調べる。ここで、キャッシュメモリ 230 に存在すれば、チャンネルコントローラ 210 はそのデータをホストの情報処理装置 400 に送信する。一方、読み出し対象となるデータがキャッシュメモリ 230 に存在しない場合には、当該チャンネルコントローラ 210 は読み出しコマンドを共有メモリ 220 に書き込むと共に、共有メモリ 220 を監視する。読み出しコマンドが共有メモリ 220 に書き込まれたことを検出したディスクコントローラ 240 は、ディスク駆動装置 300 内のディスクドライブ 310 から読み出し対象となるデータを読み出して、これをキャッシュメモリ 230 に書き込むと共に、その旨を共有メモリ 220 に書き込む。そして、チャンネルコントローラ 210 は、読み出し対象となるデータがキャッシュメモリ 230 に書き込まれたことを検出すると、そのデータをホストの情報処理装置 400 に送信する。

#### 【0035】

このように、チャンネルコントローラ 210 およびディスクコントローラ 240 の間では

、キャッシュメモリ 2 3 0 を介してデータの授受が行われ、キャッシュメモリ 2 3 0 には、ディスクドライブ 3 1 0 に記憶されるデータのうち、チャンネルコントローラ 2 1 0 やディスクコントローラ 2 4 0 により読み書きされるデータが記憶される。

#### 【0036】

なお、チャンネルコントローラ 2 1 0 からディスクコントローラ 2 4 0 に対するデータの書き込みや読み出しの指示を、共有メモリ 2 2 0 を介在させて間接的に行う構成の他、たとえばチャンネルコントローラ 2 1 0 からディスクコントローラ 2 4 0 に対してデータの書き込みや読み出しの指示を共有メモリ 2 2 0 を介さずに直接に行う構成とすることもできる。また、チャンネルコントローラ 2 1 0 にディスクコントローラ 2 4 0 の機能を持たせて、データ入出力制御部とすることもできる。

#### 【0037】

ディスクコントローラ 2 4 0 は、データを記憶する複数のディスクドライブ 3 1 0 と通信可能に接続され、ディスク駆動装置 3 0 0 の制御を行う。たとえば、上述のように、チャンネルコントローラ 2 1 0 がホストの情報処理装置 4 0 0 から受信したデータ入出力要求に応じて、ディスクドライブ 3 1 0 に対してデータの読み書きを行う。

#### 【0038】

各ディスクコントローラ 2 4 0 は、サービスプロセッサ 2 5 0 と共に内部 LAN 2 6 1 で接続されており、相互に通信を行うことが可能である。これにより、ディスクコントローラ 2 4 0 に実行させるマイクロプログラムなどをサービスプロセッサ 2 5 0 から送信し、インストールすることが可能となっている。

#### 【0039】

本実施の形態においては、共有メモリ 2 2 0 およびキャッシュメモリ 2 3 0 がチャンネルコントローラ 2 1 0 およびディスクコントローラ 2 4 0 に対して独立に設けられていることについて記載したが、これに限られるものではなく、共有メモリ 2 2 0 またはキャッシュメモリ 2 3 0 がチャンネルコントローラ 2 1 0 およびディスクコントローラ 2 4 0 の各々に分散されて設けられることも好ましい。この場合、スイッチ 2 6 0 は、分散された共有メモリ 2 2 0 またはキャッシュメモリ 2 3 0 を有するチャンネルコントローラ 2 1 0 およびディスクコントローラ 2 4 0 を相互に接続させることになる。

#### 【0040】

また、チャンネルコントローラ 2 1 0、ディスクコントローラ 2 4 0、スイッチ 2 6 0、共有メモリ 2 2 0、キャッシュメモリ 2 3 0 の少なくともいずれかが一体として構成されているようにすることもできる。

#### 【0041】

サービスプロセッサ 2 5 0 は、ディスクアレイ装置 1 0 0 を保守・管理するためのコンピュータである。オペレータは、サービスプロセッサ 2 5 0 を操作することにより、たとえばディスク駆動装置 3 0 0 内のディスクドライブ 3 1 0 の構成の設定や、ホストの情報処理装置 4 0 0、管理クライアントの情報処理装置 5 0 0 とチャンネルコントローラ 2 1 0 との間の通信路であるパスの設定、論理ボリュームの設定、チャンネルコントローラ 2 1 0 やディスクコントローラ 2 4 0 において実行されるマイクロプログラムのインストールなどを行うことができる。ここで、ディスク駆動装置 3 0 0 内のディスクドライブ 3 1 0 の構成の設定としては、たとえばディスクドライブ 3 1 0 の増設や減設、RAID 構成の変更 (RAID 1 から RAID 5 への変更など) などとすることができる。

#### 【0042】

さらに、サービスプロセッサ 2 5 0 からは、ディスクアレイ装置 1 0 0 の動作状態の確認や故障部位の特定、チャンネルコントローラ 2 1 0 で実行されるオペレーティングシステムのインストールなどの作業を行うことができる。これらの設定や制御は、サービスプロセッサ 2 5 0 が備えるユーザインタフェース、あるいはサービスプロセッサ 2 5 0 で動作する Web サーバにより提供される Web ページを表示する管理クライアントの情報処理装置 5 0 0 のユーザインタフェースからオペレータなどにより行うようにすることができる。オペレータなどは、サービスプロセッサ 2 5 0 を操作して障害監視する対象や内容の

設定、障害通知先の設定などを行うこともできる。

#### 【0043】

サービスプロセッサ250は、ディスク制御装置200に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。また、サービスプロセッサ250は、ディスク制御装置200およびディスク駆動装置300の保守・管理を専用に行うコンピュータとすることもできるし、汎用のコンピュータに保守・管理機能を持たせたものとすることもできる。

#### 【0044】

＜ディスクアレイ装置の外観構成例＞

図2および図3により、本発明の一実施の形態に係るディスクアレイ装置の外観構成の一例を説明する。それぞれ、図2は本実施の形態に係るディスクアレイ装置の外観構成を示す図、図3はディスク制御装置の外観構成を示す図、である。

#### 【0045】

図2に示すように、本実施の形態に係るディスクアレイ装置100は、ディスク制御装置200およびディスク駆動装置300がそれぞれの筐体に納められた形態をしている。図2に示す例では、ディスク制御装置200の筐体の両側にディスク駆動装置300の筐体が配置されている。

#### 【0046】

ディスク制御装置200は、正面中央部にサービスプロセッサ250が備えられている。サービスプロセッサ250はカバーで覆われており、図3に示すように、カバーを開けることにより、サービスプロセッサ250を使用することができる。なお、図3に示したサービスプロセッサ250は、いわゆるノート型パーソナルコンピュータの形態をしているが、どのような形態とすることも可能である。

#### 【0047】

サービスプロセッサ250の下部には、チャンネルコントローラ210やディスクコントローラ240、キャッシュメモリ230、共有メモリ220、スイッチ260を装着するためのスロットが設けられている。チャンネルコントローラ210やディスクコントローラ240、キャッシュメモリ230、共有メモリ220、スイッチ260は、回路基板を備えてボードとして構成されており、これらのボードが各スロットに装着される。各スロットには、これらのボードを装着するためのガイドレールが設けられている。ガイドレールに沿って各ボードをスロットに挿入することにより、チャンネルコントローラ210やディスクコントローラ240、キャッシュメモリ230、共有メモリ220、スイッチ260をディスク制御装置200に装着することができる。各スロットの奥手方向正面部には、各ボードをディスク制御装置200と電氣的に接続するためのコネクタが設けられている。

#### 【0048】

また、ディスク制御装置200には、チャンネルコントローラ210やディスクコントローラ240などから発生する熱を放出するためのファン270が設けられている。ファン270は、ディスク制御装置200の上面部に設けられる他、スロットの上部にも設けられている。

#### 【0049】

＜ディスクアレイ装置の具体的構成および動作原理例＞

図4～図8により、本発明の一実施の形態に係るディスクアレイ装置の具体的構成および動作原理の一例を説明する。それぞれ、図4は本実施の形態に係るディスクアレイ装置の具体的構成を示す図、図5はホストコンピュータからのリード要求時のチャンネルコントローラの動作を示すフローチャート、図6はチャンネルコントローラからのリード要求時のディスクコントローラの動作を示すフローチャート、図7はホストコンピュータからのライト要求時のチャンネルコントローラの動作を示すフローチャート、図8はチャンネルコントローラからのライト要求時のディスクコントローラの動作を示すフローチャート、である。

**【0050】**

本実施の形態に係るディスクアレイ装置100は、具体的にはディスクサブシステムに適用され、図4に示すように、ホストコンピュータ(=ホストの情報処理装置)400に接続され、ホストインタフェースを具備したチャンネルコントローラ210、共有メモリ220、キャッシュメモリ230、FC-AI(Fibre Channel Arbitrated Loop)インタフェースを例とするディスクドライブインタフェースを具備したディスクコントローラ240、サービスプロセッサ250と接続するためのサービスプロセッサインタフェース251、スイッチ260、ディスクドライブ310などから構成される。

**【0051】**

ディスクドライブ310は、図4においてはディスクドライブ311~315から構成される例を示しており、ディスクドライブ(D1)311、ディスクドライブ(D2)312、ディスクドライブ(D3)313、およびディスクドライブ(P)314はRAID構成により冗長性を有しており、ディスクドライブ(D1)~(D3)はデータを格納するためのドライブとして備えられ、ディスクドライブ(P)はパリティ情報を格納するためのドライブとして備えられ、またディスクドライブ(S)は予備のドライブとして備えられている。

**【0052】**

また、図4においては、サービスプロセッサ250が、サービスプロセッサインタフェース251を介してディスクサブシステムに外付けされている形態を示している。

**【0053】**

本実施の形態のディスクサブシステムにおいては、特に、ディスクドライブ310(311~315)は、記憶領域に対するデータの書き込みまたは読み出し処理においてエラーが生じた場合、データの書き込みまたは読み出しが正常に終了したことを通知した後に、データの書き込みまたは読み出しを再度繰り返し、データの書き込みまたは読み出しを実行する機能を有している。

**【0054】**

また、ディスクコントローラ240は、複数のディスクドライブ310に対する書き込みデータまたは読み出し要求が格納される格納領域を有して、複数のディスクドライブ310に対するデータの書き込みまたは読み出しを制御し、複数のディスクドライブ310からデータの書き込みまたは読み出しが正常に終了したことを通知されるとともに、特にデータの書き込みまたは読み出しに利用され、冗長性を有してデータを格納できる論理的な記憶領域を、複数のディスクドライブ310の記憶領域を用いて生成し、論理的な記憶領域を形成する複数のディスクドライブ310に関して、複数のディスクドライブ310に対する書き込みデータまたは読み出し要求が格納される格納領域を監視し、論理的な記憶領域を形成する複数のディスクドライブ310のうちデータの書き込みまたは読み出しの繰り返し回数が多いディスクドライブを特定して、特定されたディスクドライブを閉塞させる機能を有している。

**【0055】**

さらに、チャンネルコントローラ210は、自ディスクアレイ装置の外部のネットワークから書き込み要求または読み出し要求を受ける機能を有している。共有メモリ220は、チャンネルコントローラ210、ディスクコントローラ240によって通信される制御情報が格納される領域を有している。キャッシュメモリ230は、チャンネルコントローラ210とディスクコントローラ240との間で通信されるデータが一時的に保存される領域を有している。スイッチ260は、チャンネルコントローラ210、ディスクコントローラ240、共有メモリ220およびキャッシュメモリ230に接続される。

**【0056】**

また、サービスプロセッサ250は、冗長性を有する複数のディスクドライブ310のうちデータの書き込みまたは読み出しの繰り返し回数が多いディスクドライブを特定する際の条件を設定する機能を有している。

**【0057】**

このように構成されるディスクサブシステムにおける動作は、以下の通りである。ホストコンピュータ400からライトデータを受領したチャンネルコントローラ210は、キャッシュメモリ230に退避すると共にディスクコントローラ240に対して、キャッシュメモリ230にあるライトデータをディスクドライブ311～314に書き込むように指示する。また、ホストコンピュータ400からデータリード要求を受領したチャンネルコントローラ210は、ディスクコントローラ240に対し、ディスクドライブ311～314よりデータを読み出し、キャッシュメモリ230に転送するように指示する。指示を受けたディスクコントローラ240はディスクドライブ311～314よりデータを読み出し、キャッシュメモリ230に転送した後、チャンネルコントローラ210にデータ読み出し完了を報告する。報告を受けたチャンネルコントローラ210はデータをキャッシュメモリ230よりホストコンピュータ400に転送する。具体的には、以下の通りである。

**【0058】**

具体的に、ホストコンピュータ400からのリード要求時のチャンネルコントローラ210の動作は、図5に示すように、S1001でスイッチ260を介し、ディスクコントローラ240にリード要求を行う。次に、S1002でディスクコントローラ240の応答監視を行い、応答がなかった場合(No)はS1002へ、応答があった場合(Yes)はS1003へ移行する。S1003でキャッシュメモリ230にあるデータをスイッチ260を介して読み出し、ホストコンピュータ400に転送し、転送が完了したら続くS1004でホストコンピュータ400に完了報告を行う。

**【0059】**

チャンネルコントローラ210からのリード要求時のディスクコントローラ240の動作は、図6に示すように、S1011でディスクドライブ311～315にリード要求を行う。次に、S1012で転送開始を監視し、転送開始要求があった場合(Yes)、続くS1013でディスクドライブ311～315のデータを読み出し、スイッチ260を介してキャッシュメモリ230に転送する。続く、S1014で転送完了を監視し、転送が完了したら(Yes)、S1015でチャンネルコントローラ210にスイッチ260を介して完了報告を行う。

**【0060】**

ホストコンピュータ400からのライト要求時のチャンネルコントローラ210の動作は、図7に示すように、S1021でスイッチ260を介し、ホストコンピュータ400からのライトデータをキャッシュメモリ230に転送する。次に、S1022でディスクコントローラ240へスイッチ260を介してライト要求を行い、続くS1023でホストコンピュータ400に完了報告を行う。

**【0061】**

チャンネルコントローラ210からのライト要求時のディスクコントローラ240の動作は、図8に示すように、S1031でキャッシュメモリ230よりスイッチ260を介してライトするデータを読み出し、冗長データを作成し、この作成した冗長データをスイッチ260を介してキャッシュメモリ230内の別の領域に書き込む。次に、S1032でディスクドライブ311～315にライト要求を行い、続くS1033で転送開始を監視し、転送開始要求があった場合(Yes)、続くS1034でキャッシュメモリ230のデータをディスクドライブ311～315に転送し、続くS1035で転送完了を監視し、転送が完了したら(Yes)、処理を終了する。

**【0062】**

＜ディスクアレイ装置の入出力処理例＞

図9により、本発明の一実施の形態に係るディスクアレイ装置の入出力処理の一例を説明する。図9はディスクドライブに対する入出力処理を示す図である。

**【0063】**

ディスクアレイ装置の入出力処理において、キャッシュメモリ230、共有メモリ220、ディスクコントローラ240のローカルメモリ241にはそれぞれ、以下のような情



報が格納される。

#### 【0064】

キャッシュメモリ230には、ディスクドライブ311～313に対して書き込む未反映データ(D1)2301, (D2)2302, (D3)2303、ディスクドライブ314に書き込む冗長データ(P)2304、スベアのディスクドライブ315に書き込む再生データ(S)2305が格納される各領域が設けられている。さらに、キャッシュメモリ230には、ディスクドライブ311～315から読み出したリードデータ(D1')2311, (D2')2312, (D3')2313, (P')2314, (S')2315が格納される各領域が設けられている。

#### 【0065】

共有メモリ220には、ディスクドライブ311～315に対する未反映データ量のドライブ情報(D1)2201, (D2)2202, (D3)2203, (P)2204, (S)2205や、ディスクドライブの性能劣化検出用のパラメータ2206の情報が格納される各領域が設けられている。

#### 【0066】

ディスクコントローラ240のローカルメモリ241には、ディスクドライブ311～315に対応したキュー(D1)2411, (D2)2412, (D3)2413, (P)2414, (S)2415の情報が格納される各領域が設けられている。

#### 【0067】

図9において、ディスクドライブ311～313との入出力処理に関しては、チャンネルコントローラ210よりディスクドライブ311～313に対して書き込み指示のあったデータは、キャッシュメモリ230内のディスクドライブ311～313に対応する未反映データ2301～2303に格納され、これらのデータを元にディスクドライブ314に書き込むべき冗長データ2304を作成し、未反映データ量は共有メモリ220にドライブ情報2201～2204として更新される。

#### 【0068】

ディスクコントローラ240は、ディスクドライブ311～314に対し、データの書き込みが完了したら、キャッシュメモリ230内のディスクドライブ311～314に対応する未反映データ2301～2304を破棄し、共有メモリ220内のドライブ情報2201～2204の未反映データ量を更新する。

#### 【0069】

たとえば一例として、ディスクドライブ(D2)312に対する書き込み指示があったとき、このディスクドライブ312のドライブ情報2202のドライブ状態が使用不可能であった場合は、冗長データ2304を作成した後、ディスクドライブ312のドライブ情報2202に対するライト要求を行わず、未反映データ2302およびドライブ情報2202の未反映データ量を削除する。

#### 【0070】

また一例として、ディスクドライブ(D1)311に対する読み出し指示があったとき、このディスクドライブ311のドライブ情報2201のドライブ状態が使用不可能であった場合は、冗長度を有するディスクドライブ312～314よりデータを読み出し、キャッシュメモリ230上のリードデータ2312～2314に格納し、これらのデータよりディスクドライブ311のリードデータ2301を再生する。

#### 【0071】

＜キューの構成例＞

図10により、本発明の一実施の形態に係るディスクアレイ装置において、ディスクコントローラとディスクドライブで授受される入出力処理に関するキューの一例を説明する。図10はキューの構成を示す図である。

#### 【0072】

ディスクコントローラ240とディスクドライブ311～315で授受され、各ディスクドライブ311～315に対応したキュー2411～2415は、ディスクコントロー

ラ 2 4 0 のローカルメモリ 2 4 1 上に格納される。それぞれは、個数を示すキュー数 2 4 2 1 と、N 1 組の、コマンドの種別 (READ、WRITE など) を示すコマンド種別 2 4 2 2、ディスクドライブのどの位置から入出力を開始するかの位置情報を示す LBA (Logical Block Address) の入出力開始位置 2 4 2 3、入出力のデータ転送量を示すデータ転送量 2 4 2 4、入出力を要求した時間を示す要求開始時間 2 4 2 5、当該キューが使用できるか否かを示す使用情報 2 4 2 6 からなる個別キューより構成される。

#### 【0073】

たとえば、ディスクコントローラ 2 4 0 はディスクドライブ 3 1 1 ~ 3 1 5 に対し、F C - A L I / F を例とするインタフェースを介し、READ または WRITE 要求を行った際、各ディスクドライブ毎のキュー 2 4 1 1 ~ 2 4 1 5 の個別キュー (コマンド種別、入出力開始位置、データ転送量、要求開始時間、使用情報) 2 4 2 2 ~ 2 4 2 6 を登録するとともに、キュー数 2 4 2 1 をインクリメントする。また、ディスクドライブ 3 1 1 ~ 3 1 5 より READ または WRITE 要求の応答があった時に、対応する各々のキュー 2 4 1 1 ~ 2 4 1 5 の個別キュー 2 4 2 2 ~ 2 4 2 6 を削除するとともに、キュー数 2 4 2 1 をデクリメントする。

#### 【0074】

<ドライブ情報の構成例>

図 1 1 ~ 図 1 3 により、本発明の一実施の形態に係るディスクアレイ装置において、共有メモリ上に格納されるドライブ情報の一例を説明する。それぞれ、図 1 1 はドライブ情報の構成を示す図、図 1 2 はディスクドライブが使用不可能時のリード要求の動作を示すフローチャート、図 1 3 はディスクドライブが使用不可能時のライト要求の動作を示すフローチャート、である。

#### 【0075】

共有メモリ 2 2 0 上に格納されるドライブ情報 2 2 0 1 ~ 2 2 0 5 には、図 1 1 に示すように、キャッシュメモリ 2 3 0 上にあって、ディスクドライブ 3 1 1 ~ 3 1 5 へのライト未反映データの量を示す未反映データ量 2 3 2 1 と、応答時間テーブルの位置を示す応答時間ポインタ 2 3 2 2 と、ディスクドライブ 3 1 1 ~ 3 1 5 に入出力を要求した要求開始時間 2 4 2 5 から実際に応答があるまでの時間を示す応答時間 (1) 2 3 3 1 ~ (m) 2 3 3 m と、応答時間 2 3 3 1 ~ 2 3 3 m の平均を示す平均応答時間 2 3 2 3 と、当該ディスクドライブが使用可能か否かを示すドライブ状態 2 3 2 4 から構成される。

#### 【0076】

たとえば一例として、冗長構成を持つディスクドライブ 3 1 1 ~ 3 1 4 のうち、ディスクドライブ (D 1) 3 1 1 が使用不可能時のリード要求の動作は、図 1 2 に示すように、S 1 0 4 1 で共有メモリ 2 2 0 上のドライブ情報 2 2 0 1 の使用情報 2 4 2 6 を参照し、ディスクドライブ 3 1 1 が使用不可能かどうかを判断する。使用可能な場合 (No)、S 1 0 4 2 でディスクドライブ (D 1) 3 1 1 よりデータリードを行い、読み出したデータをキャッシュメモリ 2 3 0 上のリードデータ (D 1') 2 3 1 1 の領域に格納する。S 1 0 4 1 で使用不可能と判断した場合 (Yes) は、S 1 0 4 3 で冗長構成を持つ他のディスクドライブ (D 2) 3 1 2, (D 3) 3 1 3, (P) 3 1 4 よりデータリードを行い、キャッシュメモリ 2 3 0 上のリードデータ (D 2') 2 3 1 2, (D 3') 2 3 1 3, (P') 2 3 1 4 の領域にそれぞれ格納する。続く S 1 0 4 4 で、これらのデータよりディスクドライブ (D 1) 3 1 1 のデータを再生し、キャッシュメモリ 2 3 0 上のリードデータ (D 1') 2 3 1 1 に格納する。

#### 【0077】

また一例として、冗長構成を持つディスクドライブ 3 1 1 ~ 3 1 4 のうち、ディスクドライブ (D 1) 3 1 1 が使用不可能時のライト要求の動作は、図 1 3 に示すように、S 1 0 5 1 で共有メモリ 2 2 0 上のドライブ情報 2 2 0 1 の使用情報 2 4 2 6 を参照し、ディスクドライブ (D 1) 3 1 1 が使用不可能かどうかを判断する。使用可能な場合 (No)、S 1 0 5 2 でキャッシュメモリ 2 3 0 上の未反映データ (D 1) 2 3 0 1 よりデータを

読み出してディスクドライブ (D1) 311 に書き込む。S1051 で使用不可能と判断した場合 (Yes) は、処理を終了する。

#### 【0078】

＜ディスクドライブの性能劣化検出例＞

図14～図16により、本発明の一実施の形態に係るディスクアレイ装置において、ディスクドライブの性能劣化検出の一例を説明する。それぞれ、図14は共有メモリ上に格納されるディスクドライブの性能劣化検出用パラメータを示す図、図15はディスクコントローラとディスクドライブの入出力シーケンスの動作を示すフローチャート、図16 (a)～(c)は具体的な未反映データとデータ転送量の増減の関係を示す図、である。

#### 【0079】

共有メモリ220上に格納されるディスクドライブの性能劣化検出用パラメータ2206には、図14に示すように、未反映データ量の差分倍数 (n1) 2211、キュー数の差分倍数 (n2) 2212、平均応答時間の差分倍数 (n3) 2213などがある。

#### 【0080】

ディスクコントローラ240とディスクドライブ311～315の入出力シーケンスの動作は、図15に示すように、S1061でディスクコントローラ240はディスクドライブ311～315よりデータ転送要求があるかどうかを判断する。データ転送要求がなかった場合 (No) は、S1062でチャネルコントローラ210よりディスクドライブ311～315に対する入出力指示があったかどうかをチェックする。入出力指示がなかった場合 (No) は、S1061に戻る。

#### 【0081】

S1062で入出力指示があった場合 (Yes) は、S1063で入出力を発行するディスクドライブ311～315に対応する登録キュー数M1が登録可能キュー数N1未満かどうかをチェックし、未満でなければ (No) S1061に戻り、未満であれば (Yes)、S1064でキュー数M1をインクリメント (登録) し、1～N1までの間で使用情報2426が未使用のP1番目のキュー2411～2415にコマンド種別2422、入出力開始位置2423、データ転送量2424、現在の時間、すなわち要求開始時間2425を登録し、P1番目のキューの使用情報2426を使用中とし、S1065で、登録したキュー情報を元に対応するディスクドライブ311～315にキュー番号P1、コマンド種別2422、入出力開始位置2423、データ転送量2424の情報を送信し、入出力処理を要求する。

#### 【0082】

S1061でデータ転送要求があった場合 (Yes) は、S1066で転送要求のあった情報に対応するキュー番号P1のキュー2411～2415の内容に従い、データ転送を実行する。たとえば、リード転送の場合は、ディスクドライブ311～315のデータをキャッシュメモリ230上の対応する位置の未反映データ2301～2305に転送し、ライト転送の場合は、キャッシュメモリ230に対応する位置の未反映データ2301～2305からディスクドライブ311～315に転送する。

#### 【0083】

次に、S1067でデータ転送が完了したかを確認し、未完了であれば (No) ステップ1066に戻り、完了していれば (Yes)、続くS1068でドライブ情報を更新する。すなわち、応答時間 (= [現在時間] - [キュー番号P1のキューに登録された要求時間]) を応答時間ポインタ (X) 2322の示す応答時間233xに登録し、応答時間ポインタをインクリメントし (応答時間ポインタ  $X > m$  となった場合は応答時間ポインタXを0に戻す)、応答時間2331～233mより平均応答時間2323を算出する。また、完了したデータ転送がライト転送であった場合には、未反映データ量2321より転送完了したデータ転送量2424を除算する。次に、S1069で転送完了したキュー番号P1のキューの使用情報2426を未使用 (削除) とし、登録キュー数M1をデクリメントする。次に、S1070でディスクドライブの応答遅延監視処理を実施し、S1061に戻る。

## 【0084】

具体的に、未反映データとデータ転送量の増減の関係を、たとえば一例として、ディスクドライブ (D1) 311 を例に説明する。図16において、(a) に示すように、キャッシュメモリ 230 上のディスクドライブ (D1) 311 の未反映データ DD01, DD02 が存在したとすると、共有メモリ 220 上のドライブ情報 (D1) 2201 の未反映データ量 2321 は DD01 のサイズと DD02 のサイズの和となる。次に、(b) に示すように、新たなライト要求が来て、未反映データ DD03 がキャッシュメモリ 230 上に追加された場合、未反映データ量 2321 は DD01 のサイズと DD02 のサイズと DD03 のサイズの和となる。次に、(c) に示すように、新たな DD02 の未反映データがディスクドライブに書き込み完了し、未反映データ DD02 がキャッシュメモリ 230 よりクリアされた場合は、未反映データ量 2321 は DD01 のサイズと DD03 のサイズの和となる。

## 【0085】

ここでは、ディスクドライブ (D1) 311 について説明したが、ディスクドライブ (D2) 312, (D3) 313, (P) 314, (S) 315 に関しても同様である。つまり、ディスクドライブの性能が劣化し、ディスクドライブの書き込み完了が遅延した場合、未反映データのクリアも遅れるため、未反映データ量が多くなることとなる。

## 【0086】

＜ディスクドライブの応答遅延監視処理例＞

図17～図20により、本発明の一実施の形態に係るディスクアレイ装置において、ディスクドライブの応答遅延監視処理の一例を説明する。それぞれ、図17は未反映データ量に着目した応答遅延監視処理を示すフローチャート、図18は警告メッセージの表示画面を示す図、図19はキュー数に着目した応答遅延監視処理を示すフローチャート、図20は平均応答時間に着目した応答遅延監視処理を示すフローチャート、である。

## 【0087】

本実施の形態のディスクサブシステムは、RAID論理を用いた冗長度を有する構成になっており、たとえば前記図4では構成の3データドライブ+1パリティドライブ（以下3D+1Pと略す）の構成を示している。この構成では、通常、ホストコンピュータ400の一定のサイズ以上のライトデータは均等に3分割され、ディスクドライブ (D1) 311～(D3) 313 に書き込まれ、3分割されたデータより算出した冗長データはディスクドライブ (P1) 314 に書き込まれる。従って、1回のホストコンピュータ400からのライト要求に対するディスクドライブ311～314 に対して書き込みを行うべきデータ量は均一であり、それぞれのディスクドライブの性能が同等であれば、積算される未反映データ量も均等となる。もし、性能的に劣るディスクドライブが存在すれば、そのディスクドライブに対して、積算された未反映データ量は他のディスクドライブの未反映データ量よりも多くなることになる。

## 【0088】

このことより、未反映データ量を考慮し、この未反映データ量に着目した応答遅延監視処理を図17により説明する。図17では、前記図4に示す構成 (3D+1P) でディスクドライブ (D2) 312 のデータ転送完了後のディスクドライブの遅延処理を例に説明する。

## 【0089】

S1071で他のディスクドライブ311, 313, 314の共有メモリ220上のドライブ情報2201, 2203, 2204の未反映データ量2321の平均Q1を求める。次に、S1072でディスクドライブ (D2) 312の未反映データ量2321とQ1 ( $Q1 \times \text{未反映データ量差分倍数}(n1)$ ) を比較し、“未反映データ量  $> Q1 \times n1$ ” が真であった場合 (Yes)、遅延障害のディスクドライブと判断し、S1073で閉塞移行処理を実施し、偽であった場合 (No) は、正常なディスクドライブと判断し、処理を終了する。

## 【0090】

また、前述の比較で、たとえば一例として、“未反映データ量 $>Q1 \times n1 \div 2$ ”が真であった場合、遅延障害が発生する可能性のあるディスクドライブと判断し、サービスプロセッサインタフェース251を介してサービスプロセッサ250に通知し、たとえば一例として図18に示すような画面上において、「ディスクドライブ(D2)性能遅延発生」のような警告メッセージをサービスプロセッサ250上に出力してもよい。

**【0091】**

また、本実施の形態のディスクサブシステムでは、通常、ホストコンピュータ400の一定のサイズ以上のライトデータは均等に分割されることは前述の通りである。均等に分割して格納されたデータは、ホストコンピュータ400からのリード要求に対し、各ディスクドライブ311~314より均等に読み出されるということは自明の理である。すなわち、各ディスクドライブ311~314に対するリード/ライト要求の要求数(キュー数)、転送量はほぼ均一となる。もし、性能的に劣るディスクドライブが存在すれば、そのディスクドライブに対する入出力要求数(キュー数)は他のディスクドライブよりも多くなることになる。

**【0092】**

このことより、キュー数を考慮し、このキュー数に着目した応答遅延監視処理を図19により説明する。図19では、前記図4に示す構成(3D+1P)でディスクドライブ(D3)313のデータ転送完了後のディスクドライブの遅延処理を例に説明する。

**【0093】**

S1081で他のディスクドライブ311, 312, 314に対するキュー2411, 2412, 2414の登録数の平均 $Q2$ を求める。次に、S1082でディスクドライブ(D3)313のキュー2413の登録数と $Q2$ ( $Q2 \times \text{キュー数差分倍数}(n2)2212$ )を比較し、“キューの登録数 $>Q2 \times n2$ ”が真であった場合(Yes)、遅延障害のディスクドライブと判断し、S1083で閉塞以降処理を実施し、偽であった場合(No)は、正常なディスクドライブと判断し、処理を終了する。

**【0094】**

また、前述の比較で、“キューの登録数 $>Q2 \times n2 \div 2$ ”が真であった場合、前述の未反映データ量に着目した場合と同様に、遅延障害が発生する可能性のあるディスクドライブと判断し、サービスプロセッサ250に通知し、図18に示すような画面上において、「ディスクドライブ(D3)性能遅延発生」のような警告メッセージをサービスプロセッサ250上に出力してもよい。

**【0095】**

また、本実施の形態のディスクサブシステムでは、各ディスクドライブに対する入出力要求数およびデータ転送量は均等であることは前述の通りである。すなわち、ディスクドライブ311~314の性能が同等であった場合、データ転送にかかる時間(応答時間)も均等になるということは自明の理である。もし、性能的に劣るディスクドライブが存在すれば、そのディスクドライブに対する平均応答時間は他のディスクドライブよりも長くなることになる。

**【0096】**

このことより、平均応答時間を考慮し、この平均応答時間に着目した応答遅延監視処理を図20により説明する。図20では、前記図4に示す構成(3D+1P)でディスクドライブ(D1)311のデータ転送完了後のディスクドライブの遅延処理を例に説明する。

**【0097】**

S1091で他のディスクドライブ312, 313, 314の共有メモリ220上のドライブ情報2202, 2203, 2204の平均応答時間2323の平均 $Q3$ を求める。次に、S1092でディスクドライブ(D1)311の平均応答時間2323と $Q3$ ( $Q3 \times \text{平均応答時間差分倍数}(n3)2213$ )を比較し、“平均応答時間 $>Q3 \times n3$ ”が真であった場合(Yes)、遅延障害のディスクドライブと判断し、S1093で閉塞以降処理を実施し、偽であった場合(No)は、正常なディスクドライブと判断し、処理

を終了する。

#### 【0098】

また、前述の比較で、“平均応答時間 $>Q3 \times n3 \div 2$ ”が真であった場合、前述の未反映データ量に着目した場合と同様に、遅延障害が発生する可能性のあるディスクドライブと判断し、サービスプロセッサ250に通知し、図18に示すような画面上において、「ディスクドライブ(D1)性能遅延発生」のような警告メッセージをサービスプロセッサ250上に出力してもよい。

#### 【0099】

＜ディスクドライブの閉塞移行処理例＞

図21により、本発明の一実施の形態に係るディスクアレイ装置において、ディスクドライブの閉塞移行処理の一例を説明する。図21はディスクドライブの閉塞移行処理を示すフローチャートである。図21では、ディスクドライブ(D2)312の閉塞移行処理を例に説明する。

#### 【0100】

S1101で閉塞移行するディスクドライブ(D2)312に対応したキュー2412を構成する使用情報2426のドライブ状態を使用不可能(変更)とし、当該ディスクドライブ312に対する入出力要求があった場合は前述の通りとする。このことより、性能劣化のディスクドライブがディスクサブシステムより排除されることになるため、このディスクドライブに対する入出力処理時間の性能劣化が抑制され、結果としてディスクサブシステムとしての性能劣化を回避できる。

#### 【0101】

次に、S1102でスペアのディスクドライブ(S)315があるかどうかを判断し、もしスペアのディスクドライブがなかった場合(No)は処理を終了する。S1102でスペアのディスクドライブ315がシステムに存在した場合(Yes)は、S1103で他のディスクドライブ311, 313, 314よりデータリードを行い、キャッシュメモリ230上のデータ2301, 2303, 2304の領域に格納する。次に、S1104でこれらのデータの冗長度よりディスクドライブ(D2)のデータ2305を再生する。続く、S1105で再生したデータ2305をスペアのディスクドライブ315に書き込む。

#### 【0102】

次に、S1106で全てのデータを作成したかどうかを判断し、作成していなかった場合(No)はS1103に戻る。S1106で全てのデータを作成していた場合(Yes)は、S1107でスペアのディスクドライブ(S)315の使用情報2426のドライブ状態を「スペア」から「ディスクドライブ(D2)として使用可能」に変更し、処理を終了する。

#### 【0103】

＜ディスクドライブの性能劣化検出用パラメータの変更処理例＞

図22～図25により、本発明の一実施の形態に係るディスクアレイ装置において、ディスクドライブの性能劣化検出用パラメータの変更処理の一例を説明する。それぞれ、図22は性能遅延の検出レベルを変更するための設定画面を示す図、図23は検出レベルと各差分係数との対応関係を示す図、図24は応答時間・入出力処理数を変更するための設定画面を示す図、図25は検出レベルと実際のキュー数、リトライ回数との対応関係を示す図、である。

#### 【0104】

本実施の形態によるディスクサブシステムは、共有メモリ220上に格納されるディスクドライブの性能劣化検出用パラメータ2206である、未反映データ量の差分倍数(n1)2211、キュー数の差分倍数(n2)2212、平均応答時間の差分倍数(n3)2213の各係数を、サービスプロセッサインタフェース251を介して接続されたサービスプロセッサ250により変更できる。

#### 【0105】

たとえば、図 22 に示すようなサービスプロセッサ 250 の設定画面において、ユーザは、性能劣化の検出レベルを A (Easy)、B (Normal)、C (Hard) の 3 段階、および数値を直接入力できるカスタムの計 4 種類から選択できる。各検出レベルの差分倍数 (n1), (n2), (n3) の各係数は、たとえば一例として、図 23 に示すような値に設定されている。

#### 【0106】

これにより、ユーザは、通常は各係数が中間の値 (1.5) の B の検出レベルを選択し、より早く性能劣化のディスクドライブを検出したい場合は各係数が小さい値 (1.2) の A を、多少の性能劣化があってもディスクドライブを閉塞とさせず、ランニングコストを削減したい場合は各係数が大きい値 (2.0) の C をそれぞれ選択し、より細かい設定を行いたい場合は各係数を任意の値に設定できるカスタムを選択することにより、それぞれのニーズにあったチューニングが実施できる。

#### 【0107】

また、本実施の形態によるディスクサブシステムは、ディスクコントローラ 240 のローカルメモリ 241 上に格納されるキュー 2411~2415 の数 M1 (最大値) をサービスプロセッサ 250 により変更でき、またディスクドライブ 311~315 のリトライ回数もサービスプロセッサ 250 により変更できる。

#### 【0108】

たとえば、図 24 に示すようなサービスプロセッサ 250 の設定画面において、ユーザは、レベルを A (応答時間: Fast、入出力処理数: MIN)、B (応答時間: Normal、入出力処理数: Normal)、C (応答時間: Slow、入出力処理数: MAX) の 3 段階、および数値を直接入力できるカスタムの計 4 種類から選択できる。各レベルのキュー数 M1、リトライ回数は、たとえば一例として、図 25 に示すような値に設定されている。キュー数を多くした場合は多重性能が上がり、性能遅延障害がない場合には、トランザクション性能が稼げるが、性能遅延障害がある場合には、多重にした分、コマンドが沈み込む可能性がある。

#### 【0109】

こうした点より、ユーザは、通常は中間の値 (M1: 1、リトライ回数: 10) の B のレベルを選択し、1 コマンドの応答レスポンスを重視し、性能劣化/障害を早急に発見したい場合は小さい値 (M1: 1、リトライ回数: 5) の A を、障害がない場合の多重性能を重視したい場合は大きい値 (M1: 8、リトライ回数: 20) の C を選択し、より細かい設定を行いたい場合は任意の値に設定できるカスタムを選択することで、それぞれのニーズにあったチューニングが実施できる。

#### 【0110】

従って、本実施の形態のディスクアレイ装置 (ディスクサブシステム) によれば、以下のような効果を得ることができる。

#### 【0111】

(1) キャッシュメモリ 230 上のディスクドライブに書き込むべきホストコンピュータ 400 からの未反映データ量を共有メモリ 220 に保持し、この未反映データ量を冗長度を構成する他のディスクドライブと相対比較することにより、性能劣化のディスクドライブを特定することができる。

#### 【0112】

(2) 個々のディスクドライブの平均応答時間を共有メモリ 220 に保持し、この平均応答時間を冗長度を構成する他のディスクドライブと相対比較することにより、性能劣化のディスクドライブを特定することができる。

#### 【0113】

(3) 個々のディスクドライブに対するキュー数をディスクコントローラ 240 に保持し、このキュー数を冗長度を構成する他のディスクドライブと相対比較することにより、性能劣化のディスクドライブを特定することができる。

#### 【0114】

(4) 性能劣化のディスクドライブを特定し、このディスクドライブを閉塞させた後、冗長性を有する他のディスクドライブのデータを用いてスペアのディスクドライブにデータを復元することができる。

【0115】

(5) リトライ回数や、未反映データ量の差分倍数 ( $n1$ )、キュー数の差分倍数 ( $n2$ )、平均応答時間の差分倍数 ( $n3$ ) の各係数を変更できるとともに、ユーザが各レベルを選択することができるので、それぞれのニーズにあったチューニングを実施することができる。

【0116】

(6) 前記(1)～(5)により、ディスクドライブ自身に性能劣化を検出する機能がなくても、性能劣化しているディスクドライブを検出することができ、さらに性能劣化レベルを可変できることにより、ユーザの要求に応じたシステムの構築を実現することができる。

【0117】

以上、本発明者によってなされた発明を実施の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【図面の簡単な説明】

【0118】

【図1】 本発明の一実施の形態に係るディスクアレイ装置の全体構成を示す構成図である。

【図2】 本発明の一実施の形態に係るディスクアレイ装置の外観構成を示す図である。

【図3】 本発明の一実施の形態に係るディスクアレイ装置において、ディスク制御装置の外観構成を示す図である。

【図4】 本発明の一実施の形態に係るディスクアレイ装置の具体的構成を示す図である。

【図5】 本発明の一実施の形態に係るディスクアレイ装置において、ホストコンピュータからのリード要求時のチャンネルコントローラの動作を示すフローチャートである。

【図6】 本発明の一実施の形態に係るディスクアレイ装置において、チャンネルコントローラからのリード要求時のディスクコントローラの動作を示すフローチャートである。

【図7】 本発明の一実施の形態に係るディスクアレイ装置において、ホストコンピュータからのライト要求時のチャンネルコントローラの動作を示すフローチャートである。

【図8】 本発明の一実施の形態に係るディスクアレイ装置において、チャンネルコントローラからのライト要求時のディスクコントローラの動作を示すフローチャートである。

【図9】 本発明の一実施の形態に係るディスクアレイ装置において、ディスクドライブに対する入出力処理を示す図である。

【図10】 本発明の一実施の形態に係るディスクアレイ装置において、キューの構成を示す図である。

【図11】 本発明の一実施の形態に係るディスクアレイ装置において、ドライブ情報の構成を示す図である。

【図12】 本発明の一実施の形態に係るディスクアレイ装置において、ディスクドライブが使用不可能時のリード要求の動作を示すフローチャートである。

【図13】 本発明の一実施の形態に係るディスクアレイ装置において、ディスクドライブが使用不可能時のライト要求の動作を示すフローチャートである。

【図14】 本発明の一実施の形態に係るディスクアレイ装置において、共有メモリ上



に格納されるディスクドライブの性能劣化検出用パラメータを示す図である。

【図 15】本発明の一実施の形態に係るディスクアレイ装置において、ディスクコントローラとディスクドライブの入出力シーケンスの動作を示すフローチャートである。

【図 16】(a)～(c)は本発明の一実施の形態に係るディスクアレイ装置において、具体的な未反映データとデータ転送量の増減の関係を示す図である。

【図 17】本発明の一実施の形態に係るディスクアレイ装置において、未反映データ量に着目した応答遅延監視処理を示すフローチャートである。

【図 18】本発明の一実施の形態に係るディスクアレイ装置において、警告メッセージの表示画面を示す図である。

【図 19】本発明の一実施の形態に係るディスクアレイ装置において、キュー数に着目した応答遅延監視処理を示すフローチャートである。

【図 20】本発明の一実施の形態に係るディスクアレイ装置において、平均応答時間に着目した応答遅延監視処理を示すフローチャートである。

【図 21】本発明の一実施の形態に係るディスクアレイ装置において、ディスクドライブの閉塞移行処理を示すフローチャートである。

【図 22】本発明の一実施の形態に係るディスクアレイ装置において、性能遅延の検出レベルを変更するための設定画面を示す図である。

【図 23】本発明の一実施の形態に係るディスクアレイ装置において、検出レベルと各差分係数との対応関係を示す図である。

【図 24】本発明の一実施の形態に係るディスクアレイ装置において、応答時間・入出力処理数を変更するための設定画面を示す図である。

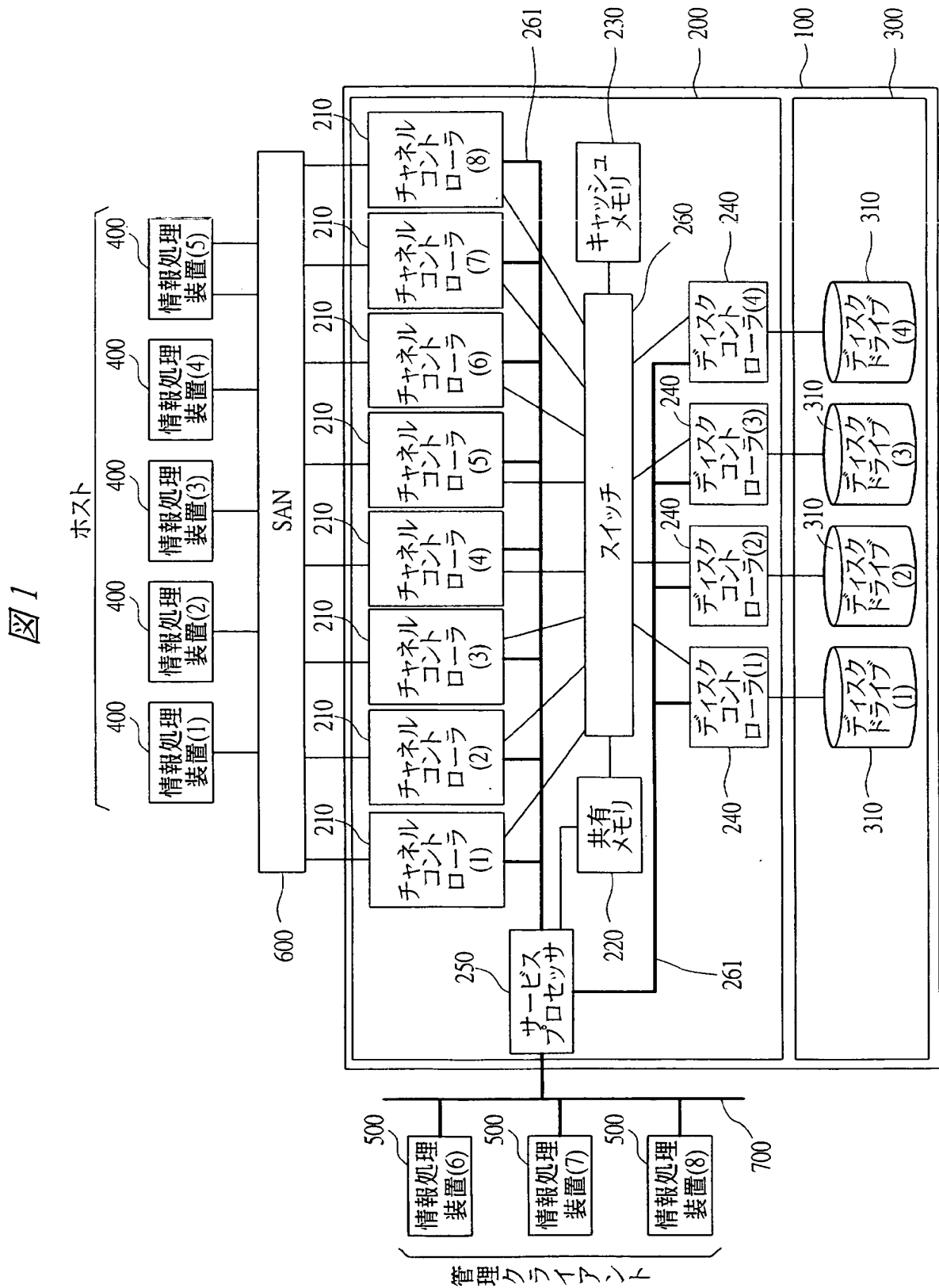
【図 25】本発明の一実施の形態に係るディスクアレイ装置において、検出レベルと実際のキュー数、リトライ回数との対応関係を示す図である。

#### 【符号の説明】

##### 【0119】

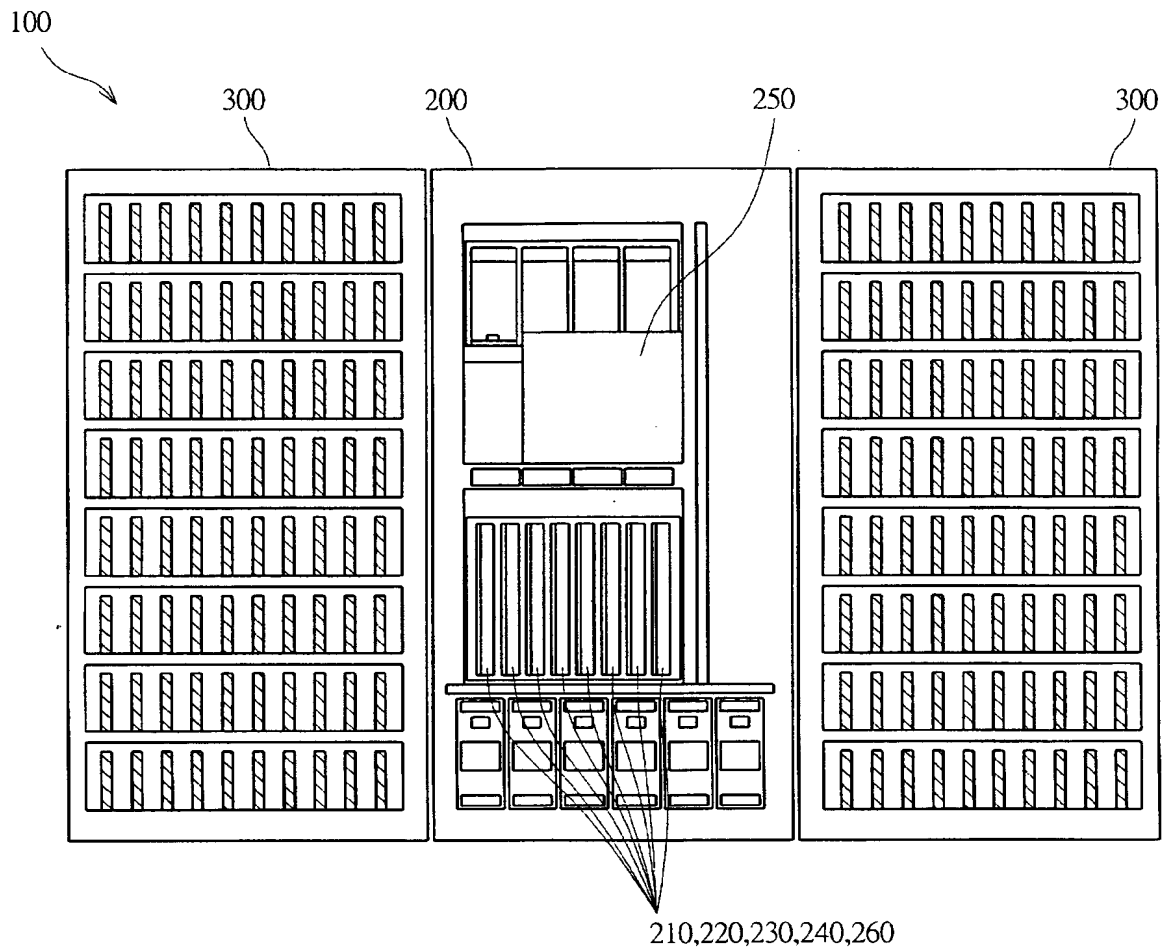
100…ディスクアレイ装置、200…ディスク制御装置、210…チャネルコントローラ、220…共有メモリ、230…キャッシュメモリ、240…ディスクコントローラ、241…ローカルメモリ、250…サービスプロセッサ、251…サービスプロセッサインタフェース、260…スイッチ、261…内部LAN、270…ファン、300…ディスク駆動装置、310, 311～315…ディスクドライブ、400…情報処理装置（ホストコンピュータ）、500…情報処理装置（管理クライアント）、600…SAN、700…LAN。

【書類名】 図面  
【図 1】



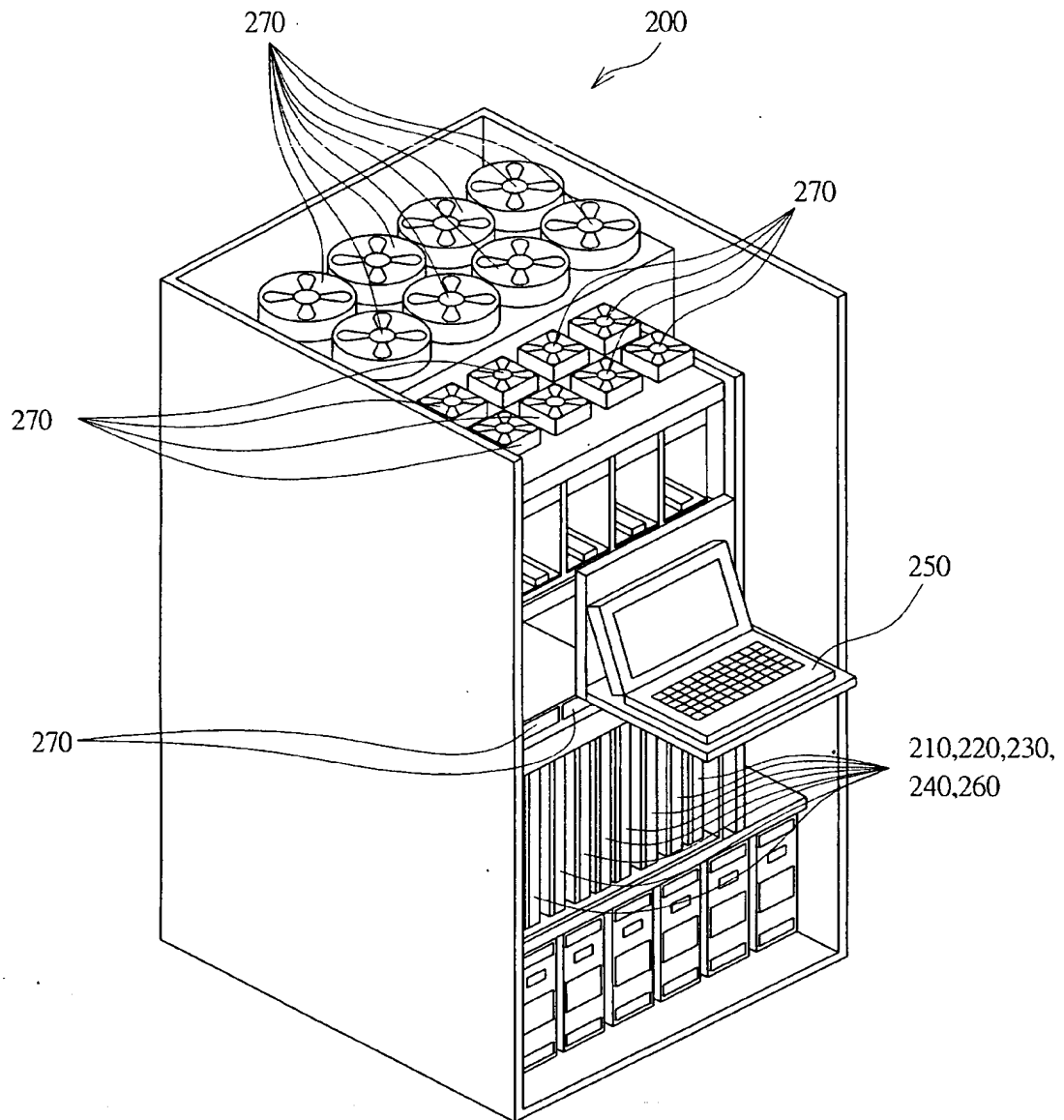
【図 2】

図 2



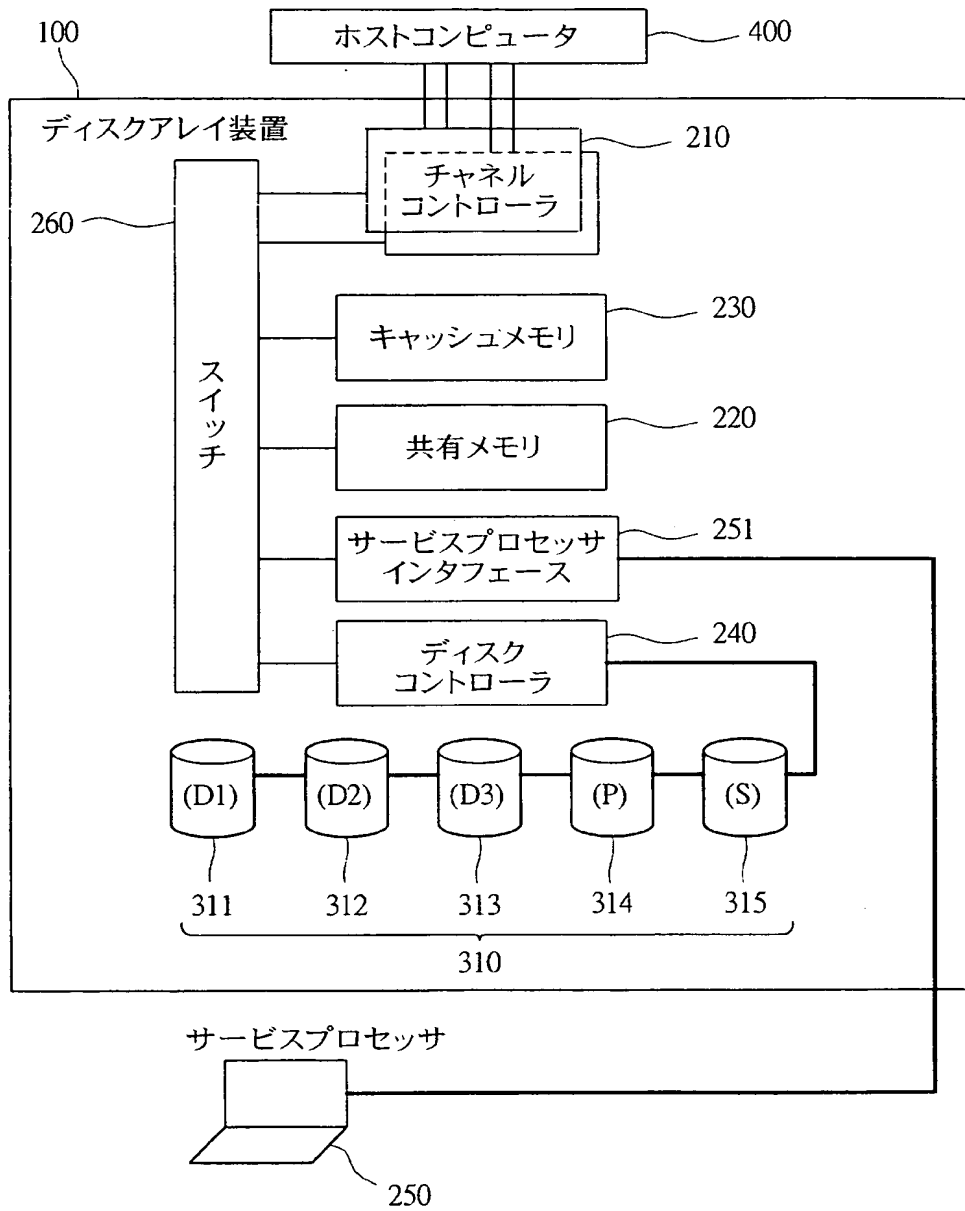
【図 3】

図 3



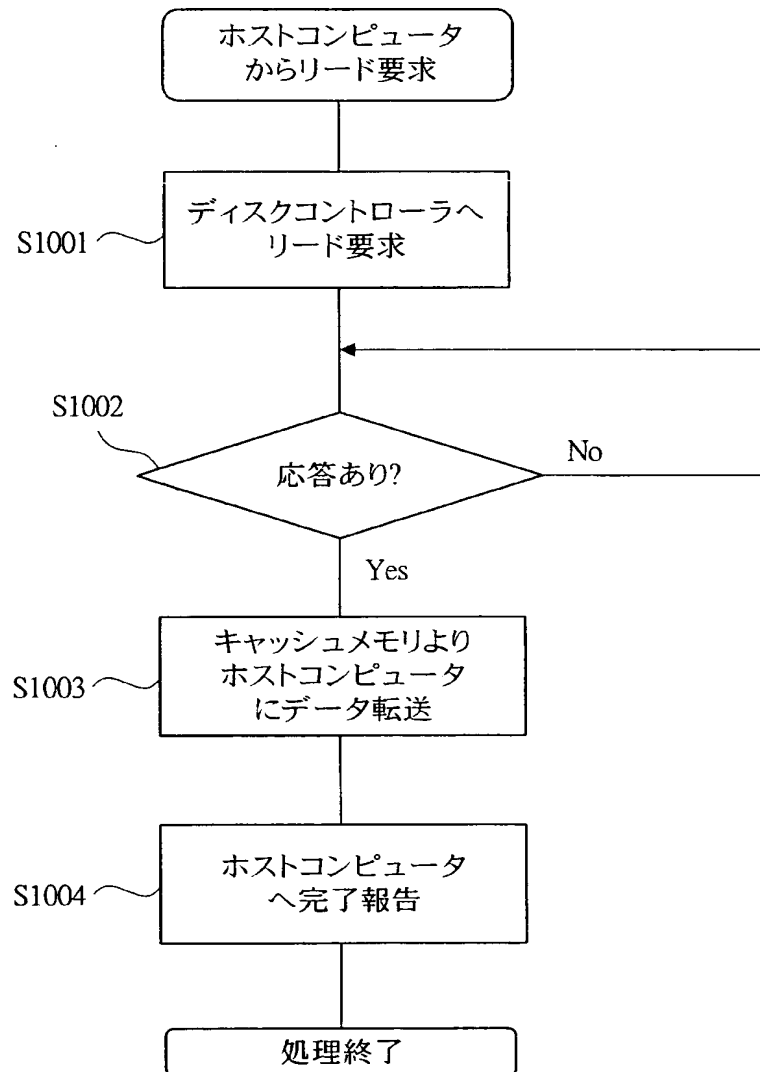
【図 4】

図 4



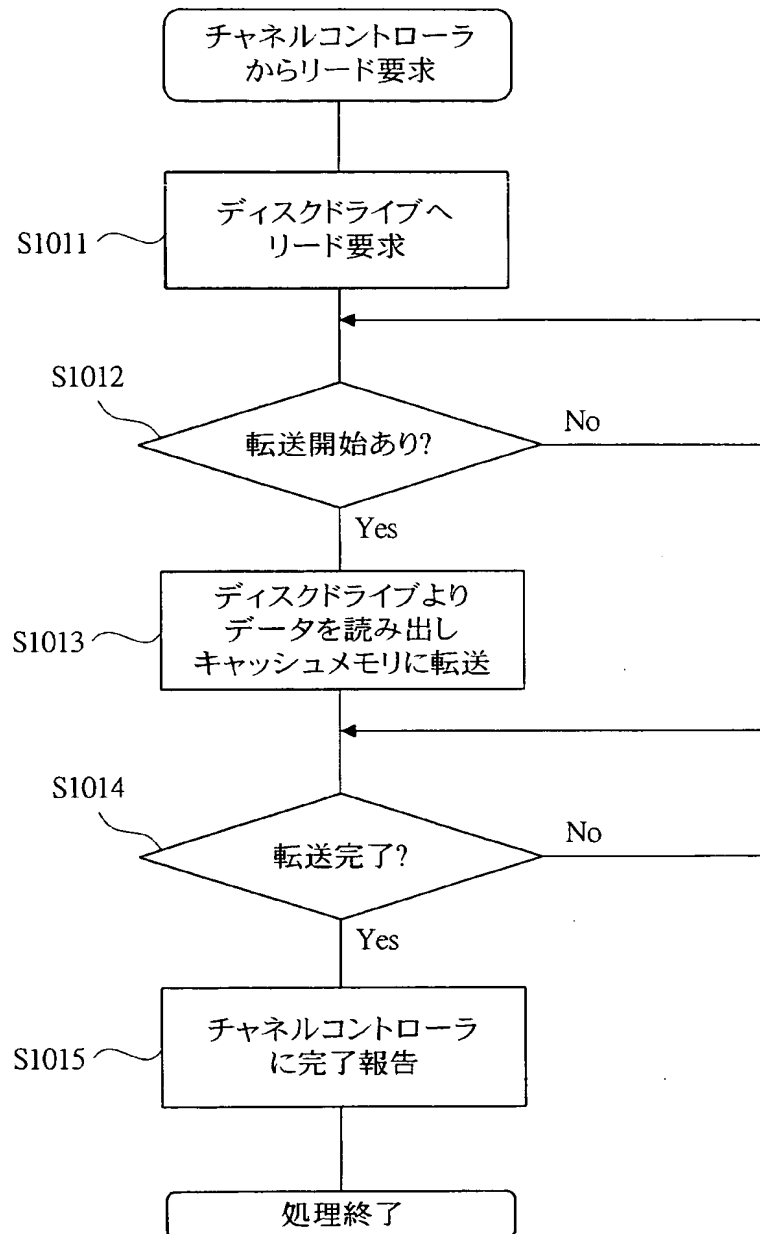
【図 5】

図 5



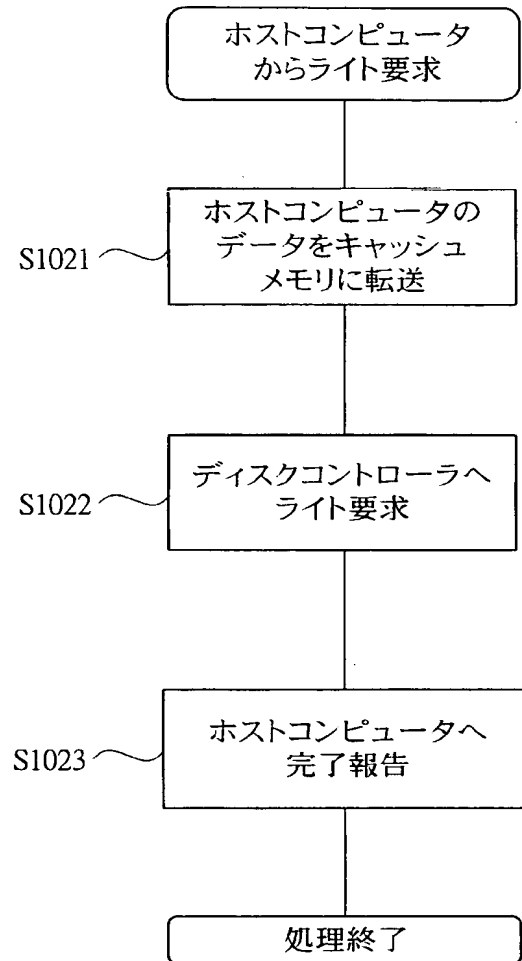
【図 6】

図 6



【図 7】

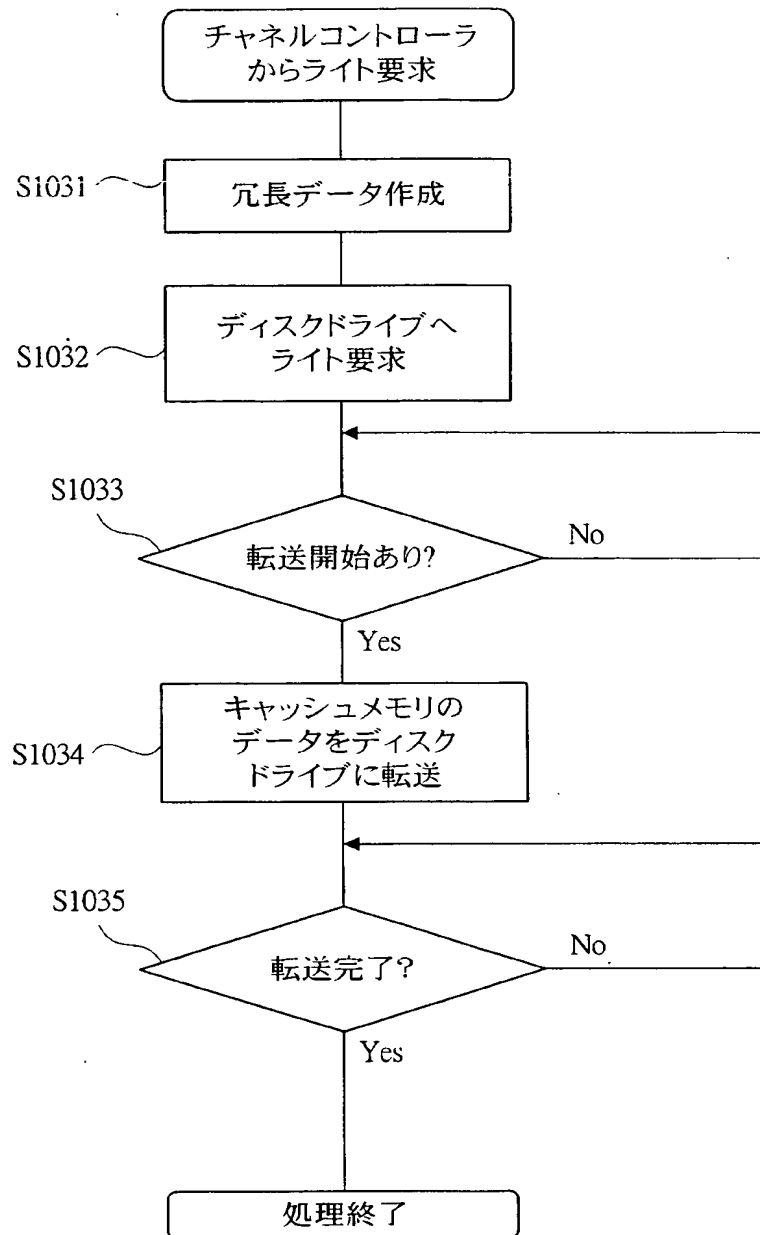
図 7





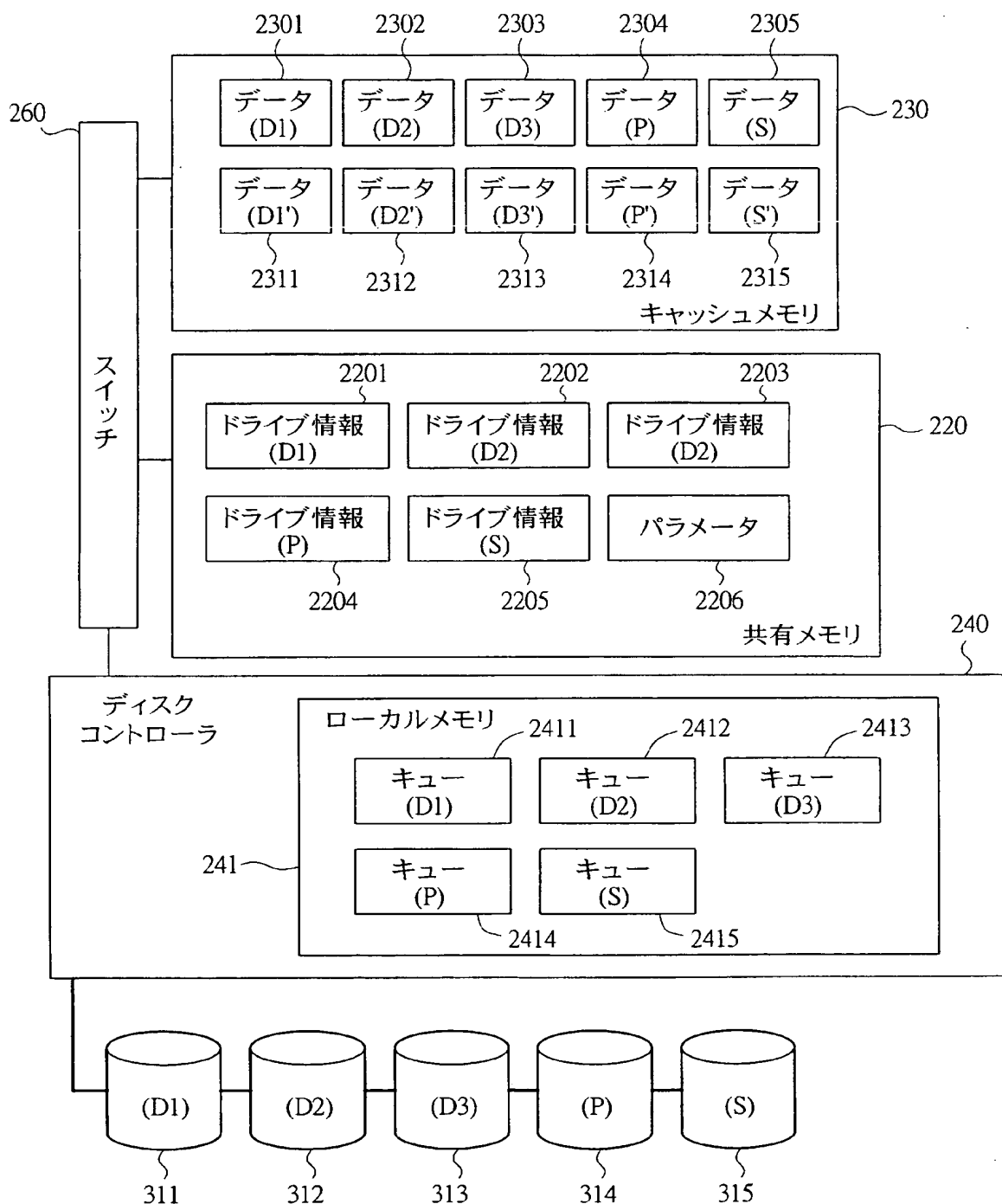
【図 8】

図 8



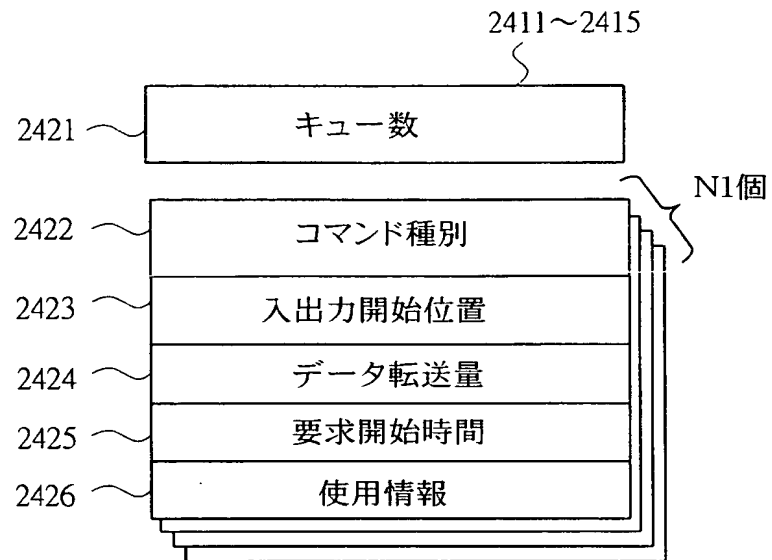
【図 9】

図 9



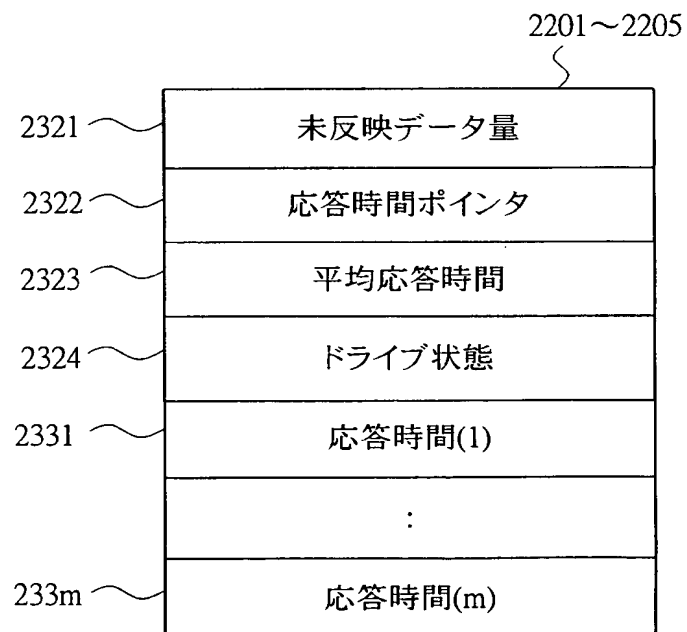
【図 10】

図 10



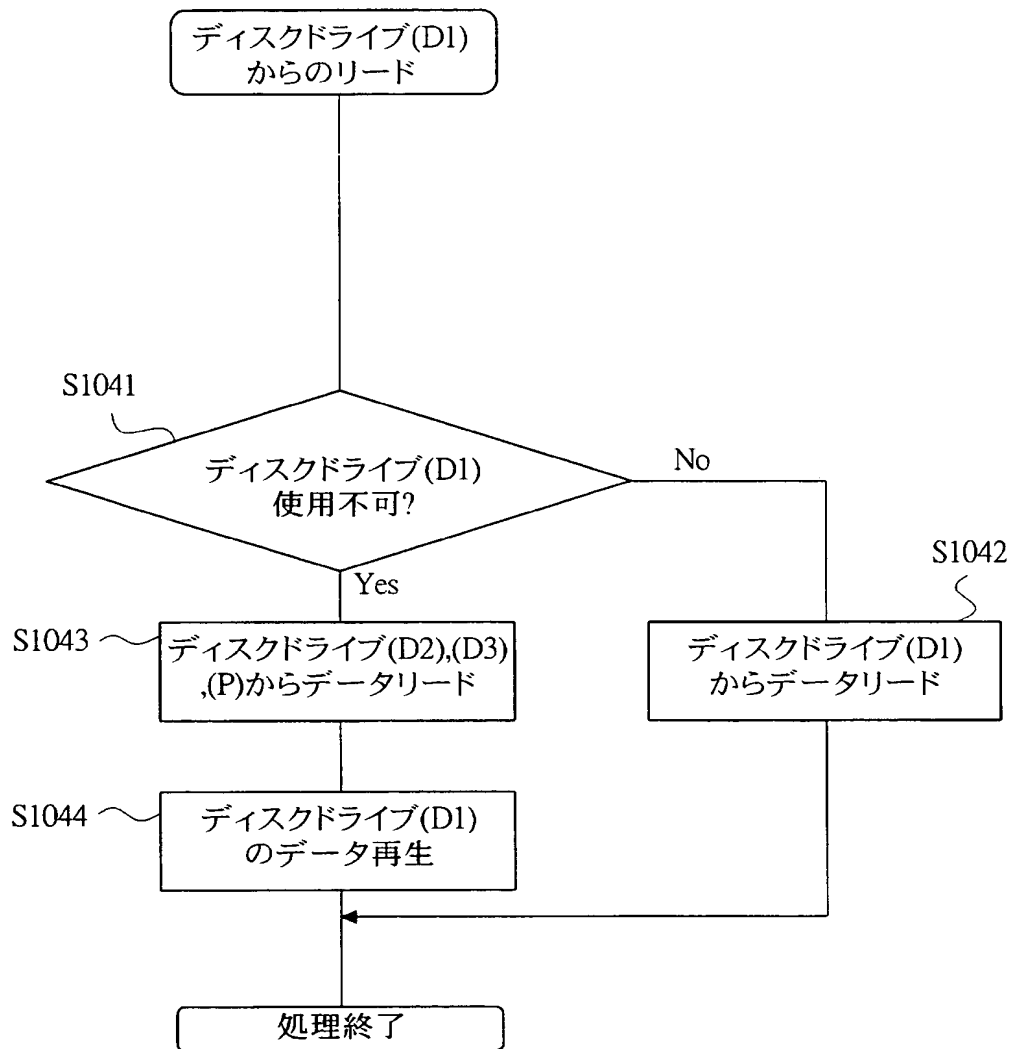
【図 11】

図 11



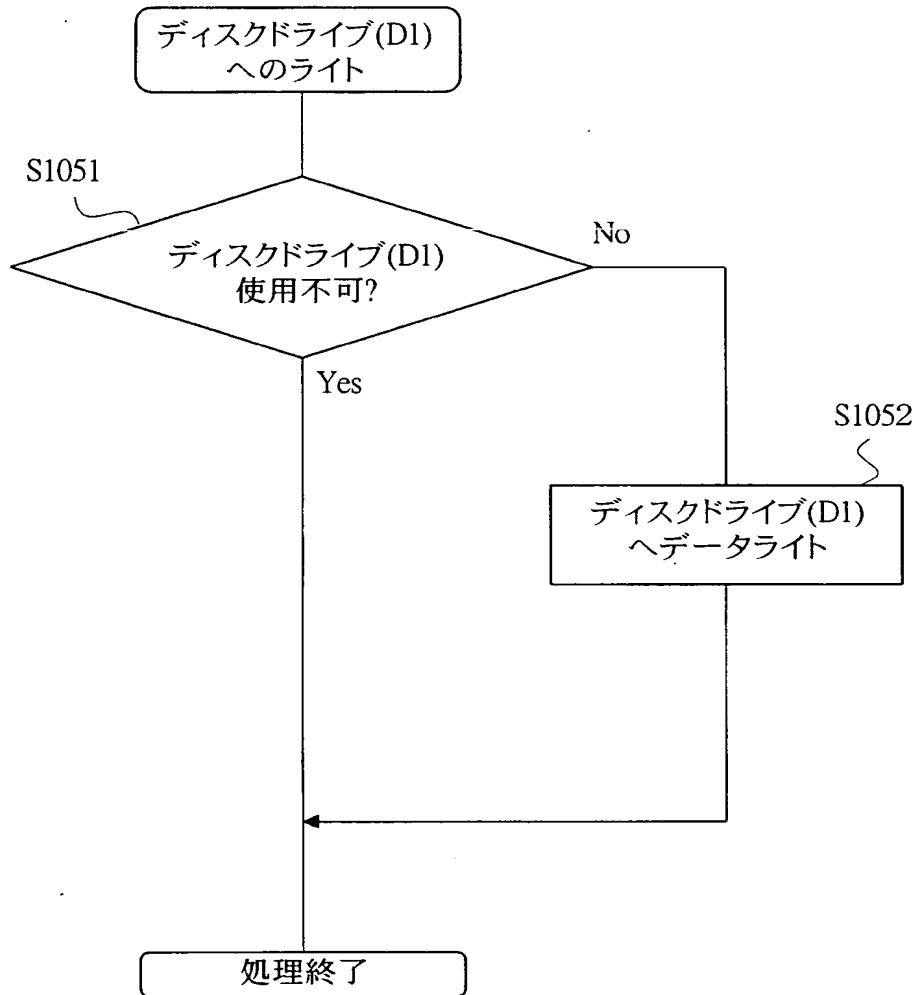
【図 12】

図 12



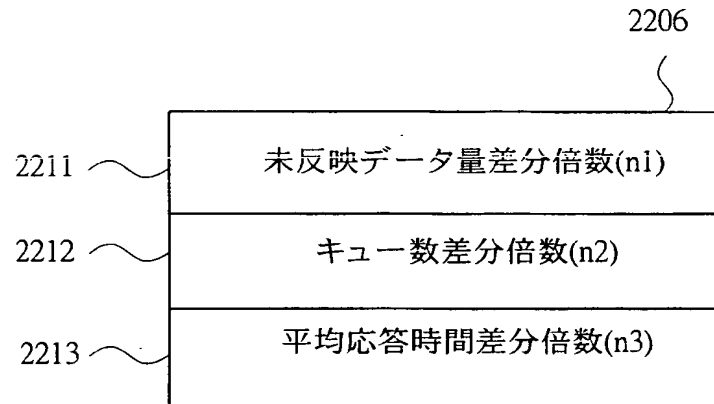
【図 13】

図 13



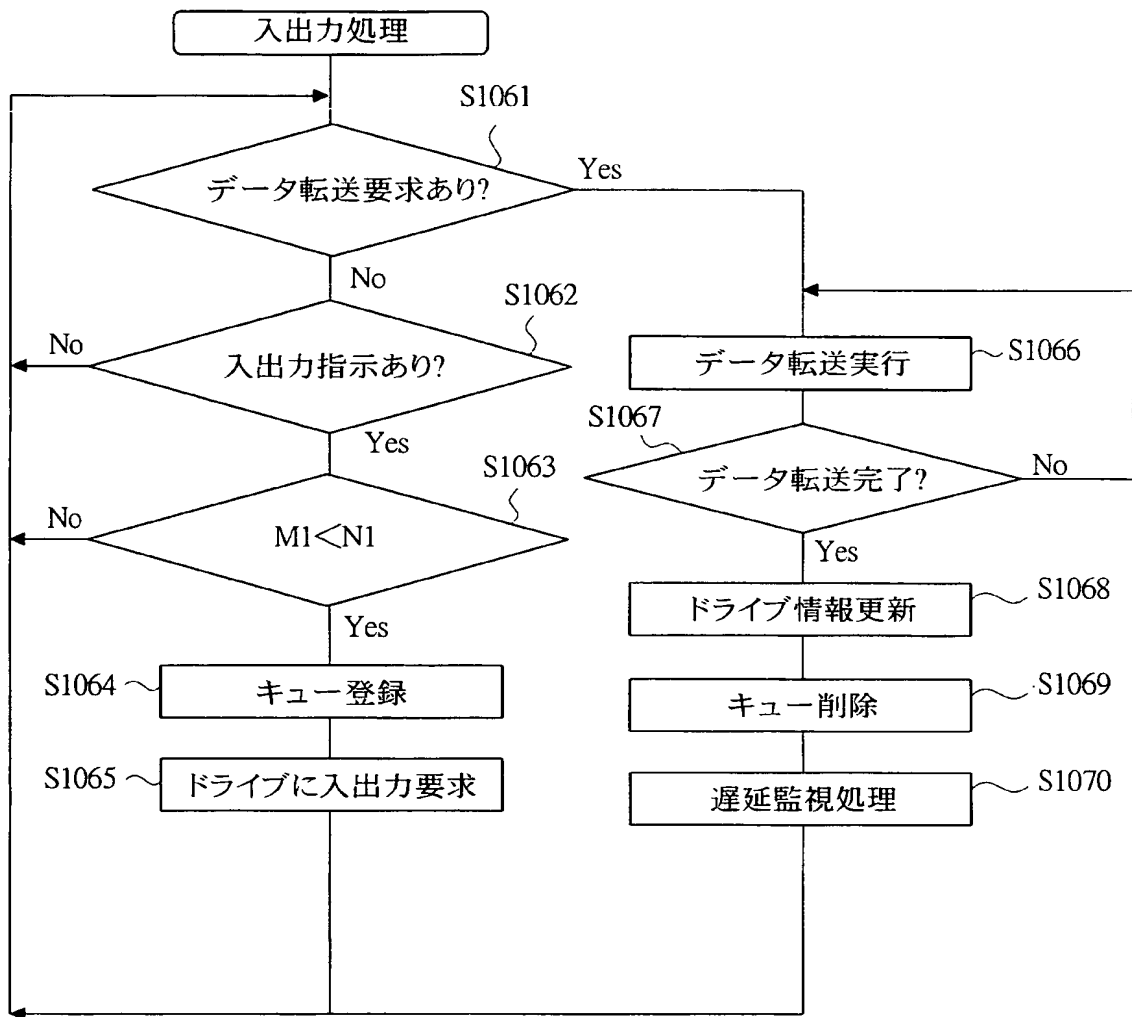
【図 14】

図 14



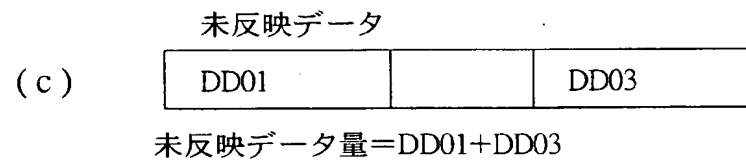
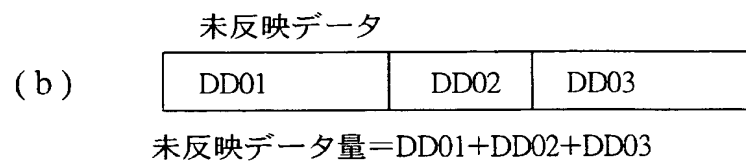
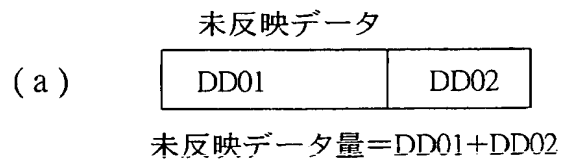
【図 15】

図 15



【図 16】

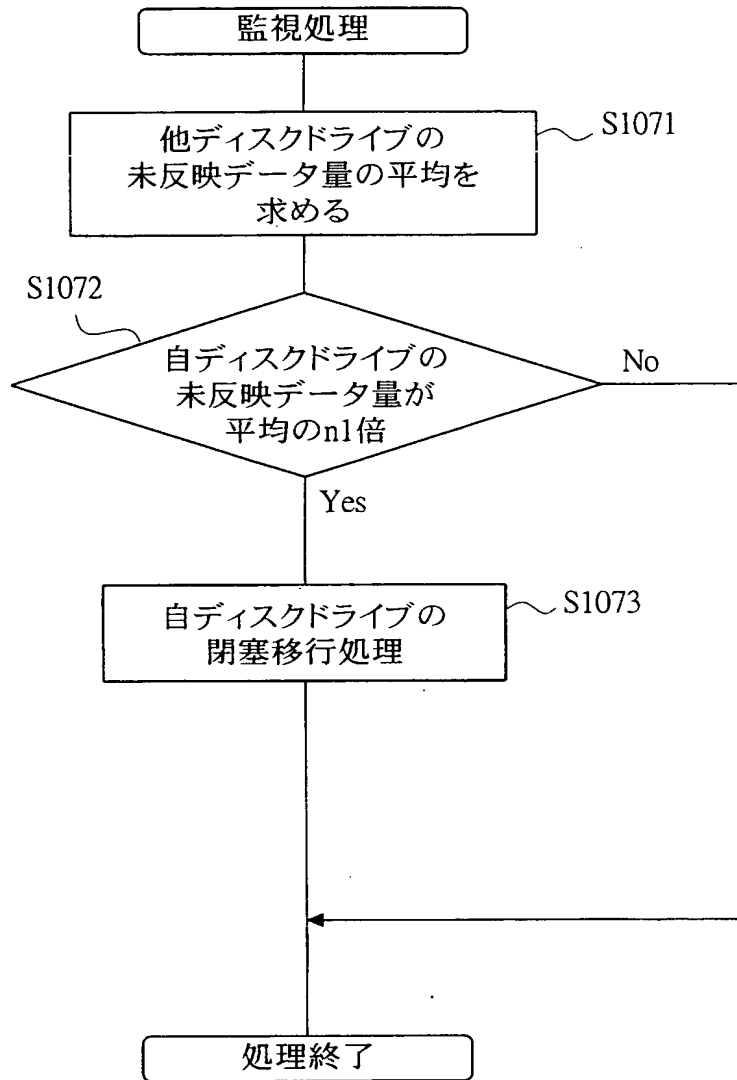
図 16





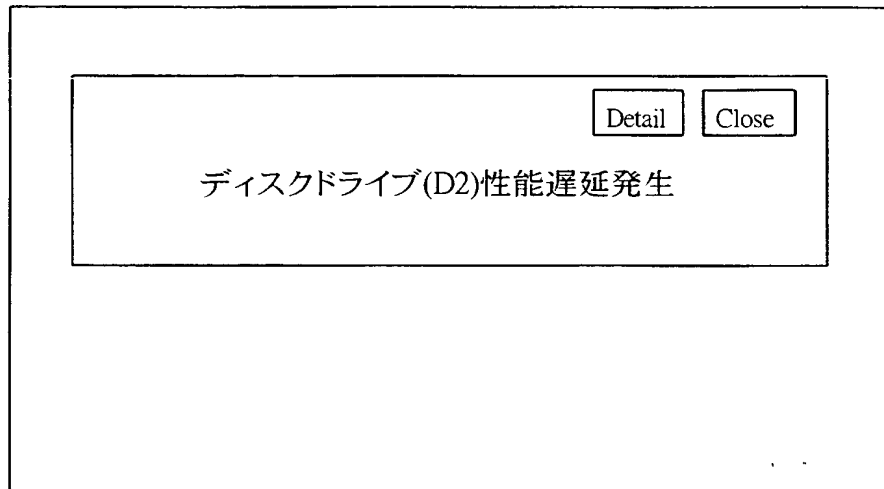
【図 17】

図 17



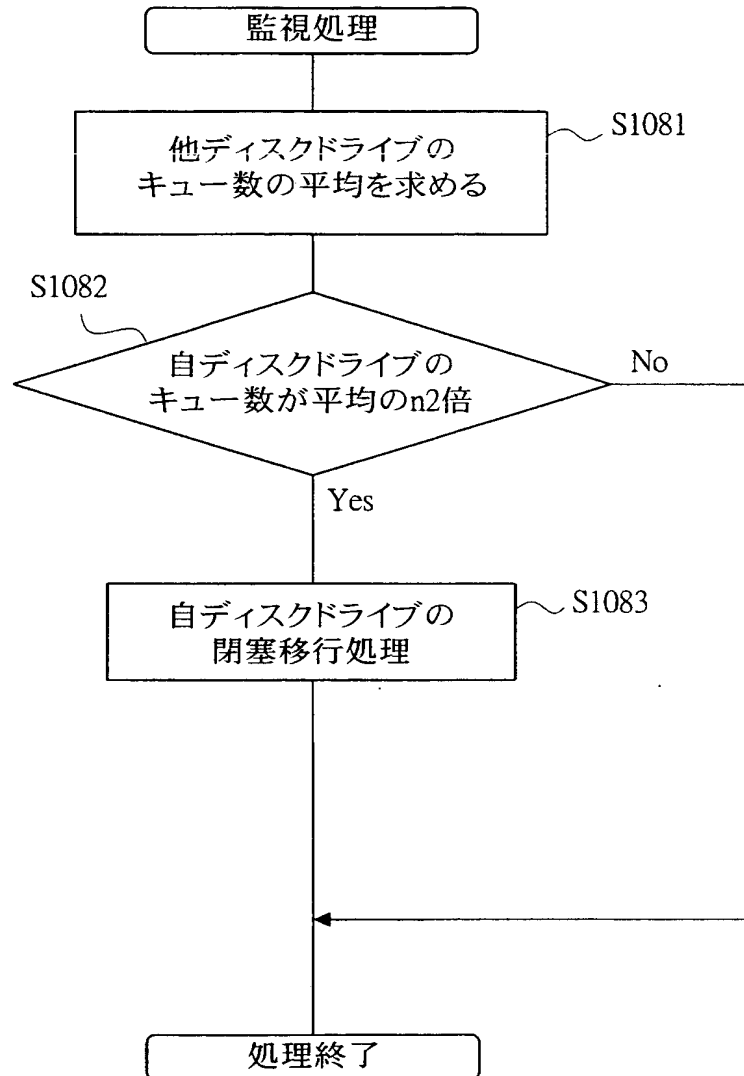
【図 18】

図 18



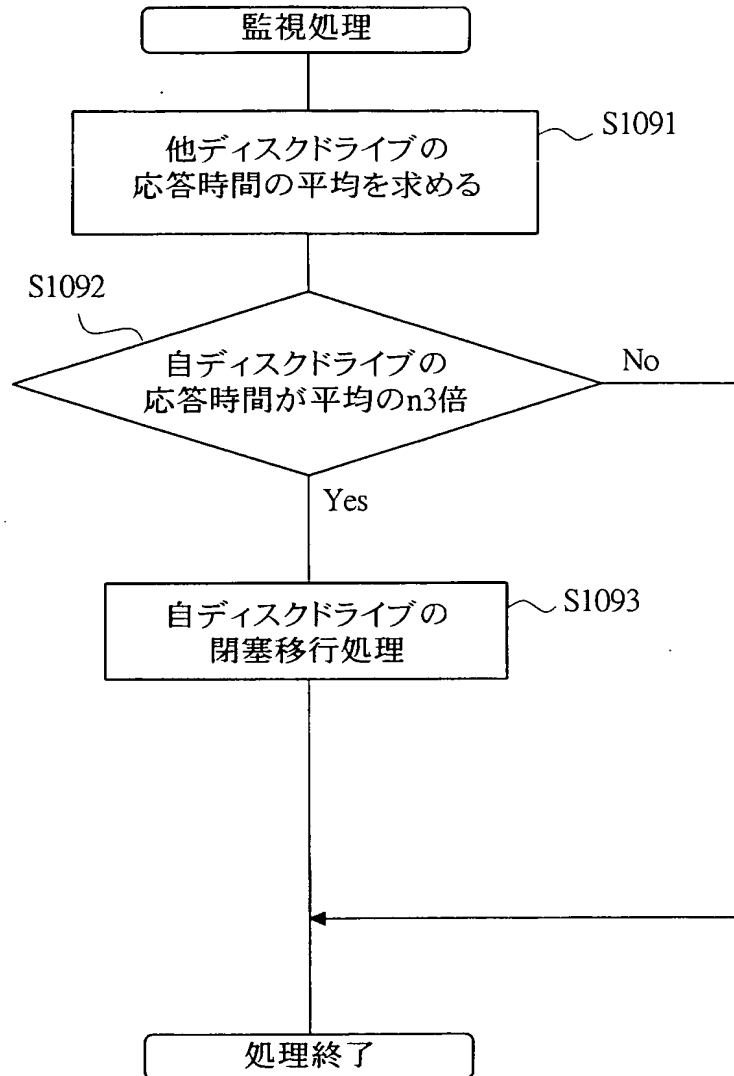
【図 19】

図 19



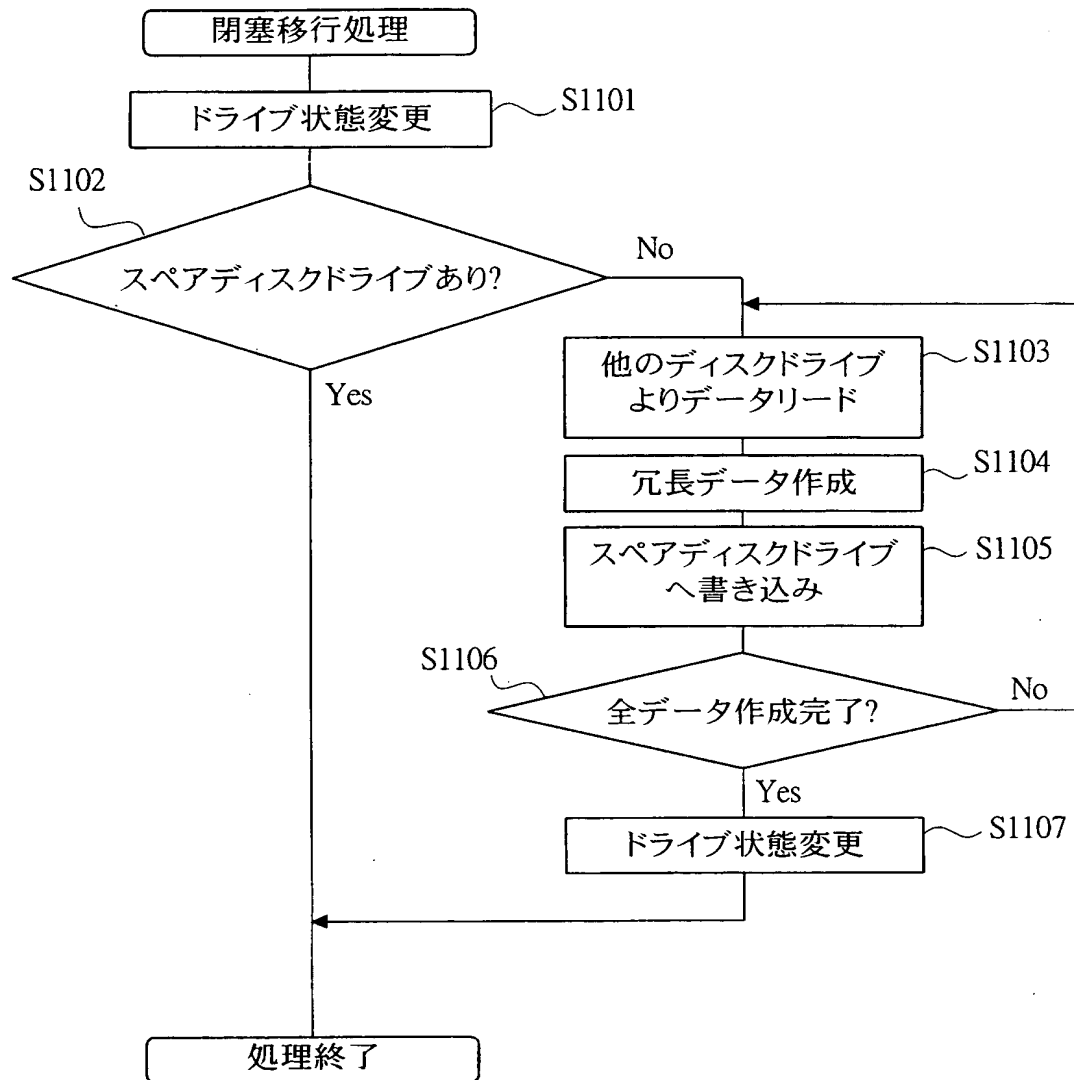
【図 20】

図 20



【図 21】

図 21



【図 22】

図 22

性能遅延検出レベル

○ A : Easy  
 ● B : Normal  
 ○ C : Hard  
 ○ Custom

未反映データ係数(1.0~4.0)   
 キュー比較係数(1.0~4.0)   
 平均応答時間係数(1.0~4.0)

【図 23】

図 23

レベル	n1	n2	n3
A	1.2	1.2	1.2
B	1.5	1.5	1.5
C	2.0	2.0	2.0
カスタム	任意	任意	任意

【図 24】

図 24

ディスク応答時間

☐ A : 応答時間

Fast

入出力処理数

MIN

☒ B : 応答時間

Normal

入出力処理数

Normal

☐ C : 応答時間

Slow

入出力処理数

MAX

☐ Custom

ドライブ設定キュー数(1~32)

リトライ回数(1~16)

【図 25】

図 25

レベル	M1	リトライ回数
A	1	5
B	4	10
C	8	20
カスタム	任意	任意

**【書類名】 要約書****【要約】**

**【課題】** ディスクドライブ自身に性能劣化を検出する機能がなくても、性能劣化しているディスクドライブを検出することができ、さらに性能劣化レベルを可変できることにより、顧客要求に応じたディスクアレイ装置の構築を実現できる技術を提供する。

**【解決手段】** ディスクアレイ装置 100 において、ディスクコントローラ 240 は、データの書き込みまたは読み出しに利用され、冗長性を有してデータを格納できる論理的な記憶領域を、複数のディスクドライブの記憶領域を用いて生成し、論理的な記憶領域を形成する複数のディスクドライブに関して、複数のディスクドライブに対する書き込みデータまたは読み出し要求が格納される格納領域を監視し、論理的な記憶領域を形成する複数のディスクドライブのうちデータの書き込みまたは読み出しの繰り返し回数が多いディスクドライブを特定して、特定されたディスクドライブを閉塞させる。

**【選択図】 図 4**



特願 2 0 0 4 - 0 6 3 3 1 3

出 願 人 履 歴 情 報

識別番号 [ 0 0 0 0 0 5 1 0 8 ]

1. 変更年月日	1 9 9 0 年 8 月 3 1 日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台 4 丁目 6 番地
氏 名	株式会社日立製作所